Is Your Peer a Lemon?

Relative Assessment of Risk Remuneration on the P2P Lending Market

Abstract: Using a sample of 11,752 loans from the Prosper peer-to-peer lending marketplace, this study employs a 5-stage methodology in order to analyze and compare the attractiveness of the P2P lending market with traditional investment alternatives in terms of risk remuneration. Results of default probability modeling in Stage I, loan expected return calculation in Stage II and efficiency frontier construction in Stage III present the evidence of high potential of P2P lending market as an investment alternative to the stock market, assuming maximum diversification opportunities and lender efficiency in interest rate setting. The subsequent stages (IV and V), while relaxing the two aforementioned assumptions, arrive to the conclusion that at its current level of development the peer-to-peer loan market offers an attractive investment risk remuneration particularly for lenders with longer investment horizon or with lower financial literacy.

Keywords: Peer-to-peer (P2P) lending, Social Lending, Risk Remuneration, Sharpe ratio, Individual investment decision making

Author: Dmitrijs Golubnicijs*
Tutor: Paolo Sodini
Date: 2012-05-14

*40146@student.hhs.se
The author would like to thank Prof. Paolo Sodini for his valuable guidance and inputs throughout the thesis project as well as Valentina Pozzobon for continuous supply of love, support and tasty food.

Table of Contents

List of Tables

List of Figures

1 Introduction

Suppose your peer John has a difficult situation in life – the recent hurricane has destroyed the house where he lived together with his family. His savings do not allow him buying the new house straightaway, so he decided to turn to a local bank. Despite a stable job and decent joint family income, the bank refuses to issue a new loan for John because, first, he already has taken loans for a car and children education, and, secondly, the hurricane brought to the bank severe losses by destroying John's neighborhood. In this situation John has two options – take his family to the shelter or ask people like you to lend him 50,000 USD. The second option might seem attractive for both of you – by eliminating the bank as a middleman you can get relatively high return on the investment of your savings and John can enjoy lower interest rates and the new accommodation. As you do not know John's family so well, you are reluctant to lend 50,000 USD collateralized only with the trust in John's job and unverified family joint income. The maximum you would be ready to lend would be 500 USD at the minimum rate of 5%. John is aware of your concerns and approaches as many other peers as possible with identical request to lend him the remaining sum of money at maximum rate of 10%. By being able to build trust through disclosing a lot of information about the difficult situation he is in, the financial status of his family members and recommendations from his employer, John managed to persuade 110 people to lend him 500 USD each with the minimum borrowing rate ranging from 5% to 20%. John ranks all borrowers by the interest rate and set the borrowing rate at the 100th lowest one, satisfying 100 borrowers who were ready to pay lower rates. As a result, both parties receive what they wanted from this peer-to-peer money transaction.

The development of the internet across the world has allowed people similar to John to reach out to a greater number of peers with borrowing requests. This resulted in the emergence of Zopa, the first peer-to-peer lending marketplace, in 2005. However, the concept quickly gained wider support in the internet society only around the start of the recent financial crisis, due to growing anger towards the banking sector and the increasing need for funding/investment alternatives. Since then, the peer-to-peer lending market has sky-rocketed with two major players – Prosper and LendingClub – providing funding to more than 950 million USD in loans. Undoubtedly, the concept also caught the interest of the academic world analyzing disintermediation in the banking industry.

The benefits of peer-to-peer platforms to two participating parties could be summarized as follows: borrowers gain from higher probability of getting a loan funded at lower interest rates, while lenders would gain from higher expected remuneration for the investment risk relatively to other investment alternatives. Judging upon the existing literature, researchers primarily focused on determining borrower benefit maximizing factors – stronger financial situation (Ryan et al., 2007; Herzenstein et al., 2008; Ravina, 2008; etc), more soft unverifiable information like borrower's effort or loan purpose (Herzenstein et al., 2011; Pope &

Sydnor, 2011; Duarte et al., 2010; etc.) or social information about borrower's friends and groups (Lin et al., 2009; Herrero-Lopez, 2009; Freedman & Jin, 2011; etc.). However, the lenders' perspective on the market place has been relatively overlooked. Few studies only, such as Iyer et al. (2011) or Miller (2011), have analyzed whether the factors maximizing borrower benefits coincide with the ones maximizing lender benefits through lower default rates and higher investment returns. Earlier studies by Kumar (2007) and Herzenstein et al. (2008) were later contrasted by Klafft (2008), Freedman & Jin (2011) and Ceyhan et al. (2011) suggesting that peers asking for money on peer-to-peer platforms are lemons and, thus, an average investor loses money.

This study enriches the discussion by analyzing the market from the investment risk remuneration perspective and then comparing the relative attractiveness with the traditional risky investment alternative – the stock market. This thesis thus intends to understand under which conditions, if at all, the peer-to-peer lending could yield higher investment risk remuneration relatively to the stock market.

A five-stage methodology is developed for this purpose. Stage I uses the Cox proportional hazard model in order to find P2P loan default determinants and to model individual loan survival probabilities. Taking the latter into account, Stage II calculates the expected return for the lender from each potential investment considering the specifics of loan repayments on the Prosper P2P lending marketplace. Stage III combines the outputs of preceding two stages for sorting loans into loan classes via classification and regression tree analysis and consequent efficiency frontier construction for the whole marketplace. The output of Stage III provides preliminary evidence so as to answer the research question, i.e. whether P2P loan market can offer relatively higher risk remuneration than stock market investments at all in the best case scenario. The latter assumes the maximum diversification opportunities over the entire sample period and lenders' efficiency and rationality in borrowing rate setting. If the possibility of P2P loan market relative superiority in terms of risk remuneration exists, Stage IV and V seek to find the conditions for it by relaxing two aforementioned assumptions and adjusting Stage III results to actual level of the Prosper marketplace development.

Empirical findings allow rejecting the main hypothesis of the paper that P2P loan market are unattractive as an investment alternative comparing to the stock market with respect to risk remuneration. The optimal portfolio in the best case scenario on the Prosper marketplace offers a Sharpe Ratio of 3.20 instead of an average Sharpe Ratio of 1.25 for S&P 500 over the sample period, in line with the suggestion of Herzenstein et al. (2008) and in contrast with papers by Klafft (2008), Freedman & Jin (2011) and Ceyhan et al. (2011). The adjustment to actual loan availability for diversification purposes on a month-to-month basis shows that peer-to-peer loans significantly underperform relatively to the stock market during turbulent times such as the recent financial crisis. However, as gaining on short-term volatility of stock returns and, thus, extracting the aforementioned benefit of stock market investments requires investors to be highly financially literate, the investments in peer-to-peer lending become more attractive under the conditions of longer investment horizons or lower financial literacy of borrowers. The findings are robust with respect to lender inefficiency in setting borrowing rates.

The paper is structured as follows. First, the existing set of literature on peer-to-peer lending is overviewed, eventually defining the research contribution of this study. Then, building on the existing literature, the methodology used for the analysis is described in detail. The subsequent data description section is followed by empirical result discussion. Finally, concluding thoughts are presented together with suggestions for further research.

2 Literature Review on Peer-to-Peer Lending

The review of relevant theoretical underpinnings is essential to achieve a sound understanding on the peer-to-peer lending concept, while forming the starting point for the hypothesis development process and the methodology chosen to test it. Generally speaking, despite the early stage of the concept development, a broad stream of research already exists on the peer-to-peer lending market and its performance. In this section, the peer-to-peer lending concept and its positioning with respect to conventional financial intermediaries are described at first; further, risks and benefits associated with peer-to-peer lending are presented. Last, the literature review section covers both the main risks and the major benefits of social lending. By identifying information asymmetry as the main challenge behind peer-to-peer lending, the related research stream on its potential mitigation is presented. Then, the commonly recognized advantages of the P2P lending market are presented for both the borrower and the lender perspectives: the success of the former could be expressed by the probability to get the loan and to reduce the interest rate, while the success of the latter depends on the loan repayment and its link with the expected return.

2.1 Peer-to-Peer Lending Market and Market Players

The rapid development of the Internet, brought by the dot.com bubble, together with the growing Internet penetration around the world prepared a fertile breeding ground for the peer-to-peer network appearance. The advent of peer-to-peer networks gradually expanded coverage areas, moving from file sharing to a more complex exchange of financial resources. The first online peer-to-peer lending platforms, also known as person-to-person (P2P) or social lending, were pioneered by Zopa (www.zopa.com) in 2005.

Since then the market has grown rapidly; as of March 2011 Zopa has facilitated more than 200 million USD; the two biggest P2P lending platforms – Prosper (www.prosper.com) and LendingClub (www.lendingclub.com) have experienced an impressive growth – only in 2011, for example, Prosper platform grew by 179%, and, as of December 31, 2011, the total amount of outstanding loans was equal to 75.76 million USD (Prosper – Media Room, 2012). The geographical reach is growing as well. Currently peer-to-peer lending platforms are known to have a solid presence in lending markets of US, UK, Germany, France, Japan, China, South Korea, India and Australia (United States Government Accountability Office, 2011).

It is crucial to recognize that the P2P lending market is "somewhat of a misnomer" (Gallloway 2009), because, in fact, no platform allows lenders to lend directly to borrowers. From a legal standpoint, platforms either broker loan reimbursements through interest-free investments, or broker the sale of securities backed by their issuers, or facilitate the origination of loans which are then sold as securities to P2P investors who behave like lenders (Galloway, 2009). For the sake of consistency and clarification, in this work P2P, or social lending, will refer to all three aforementioned operational models.

As implied above, peer-to-peer platforms differ dramatically in type and approach used to reach their social lending goal. Some choose to connect two parties directly, while others prefer the usage of a third-party intermediary. The interest rate charged is also very dependent on the platform chosen: the interest rate is pre-set by some platforms based on specific algorithms and borrower characteristics, whereas in other platforms it is left to the negotiation between lenders and borrowers, as a typical result of supply-demand. As a consequence, all market players in the P2P lending business can be classified with respect to 1) the level of control exerted over funding and 2) the pricing of loans. For example, lenders at LendingClub can directly and freely contribute to a borrower's loan request, yet the interest rate that they will receive is set in advance; in other words, LendingClub is using direct funding and intermediated pricing models. Another example is the mechanism employed by the well-known microfinance platform Kiva (www.kiva.org), which uses an indirect funding model by enabling lenders to invest in already issued loans and intermediated pricing model since the return on these loans is pre-determined by the platform. Figure 1 depicts the above-identified market player segmentation, with the most representative examples of platforms.

Among different platforms, of particular interest is Prosper, one of the biggest market players in the P2P lending business. Despite Prosper shifted to intermediated pricing model in the end of 2011, this platform possesses a particular appeal to researchers due to its characteristic of giving relatively more power to participating parties with respect to funding and pricing of loans. It is this feature that both triggered and at the same time enabled the detailed analysis of the underlying individual investment decisions, and lead to majority of academic papers being based on this platform (Bachmann et al., 2011).

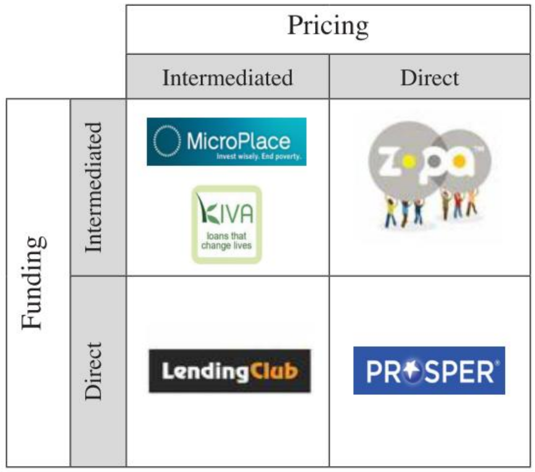

Figure 1 - P2P Market Player Segmentation

The figure depicts the peer-to-peer market segmentation with respect to the type of pricing and funding of P2P loans that each platform employs. "Direct" stands for the process without any middleman; while "Intermediated" – for the platform acting as the intermediate between borrowers and lenders.

2.1.2 Prosper and Its Operational Model

As already noted before, the Prosper platform is particularly relevant for the analysis of individual investment decisions and thereby for this research in particular. The platform was founded in February 2006, has funded 335 million USD and has 1.32 million registered members (Prosper – About Us, 2012). The platform follows the reverse Dutch auction system, where bids are expressed in interest rates at which lenders would be ready to finance a particular loan. All loans on the marketplace are uncollateralized and entail fixed repayment schedules over 3 possible maturities – one, three and five years. The business model of the platform is based on charging borrowers and lenders a fee of 1-2% and 0.5-1% of the funded amount respectively. Although there is no secondary market for Prosper loans, the platform reserves the right to sell them to debt collecting agencies in case a borrower defaults. The following example might give a good representation of the platform operation model:

> Suppose John would like to borrow 2,500 USD at the maximum rate of 15%. There are 110 lenders registered at Prosper who, based on John's reported financial and private information, believe that he will definitely repay the loan back during 3 years. Despite this belief, none of lenders knows John personally to fund the loan in full. Therefore, each investor commits to a minimum allowed sum of 25 USD specifying the minimum interest rate at which he/she would be ready to fund John's loan. Lenders get priority for the loan based on the minimum interest rate they are willing to accept, with low-rate bids getting higher priority. Those lenders left out of the loan funding can bid down the interest rate if their risk assessment of John's

loan has changed and they still would like to be entitled to 1/100th of his loan repayment over next 3 years. After a 7 day bidding period, the final interest rate for the loan is determined by the 100th lowest reservation rate.

When creating a loan request on Prosper, also called as a listing, the borrower discloses the information he/she thinks the lender might need to make a favorable investment decision. Researchers divide the possible information provided into hard, soft and social information sources.

Hard information is based on the credit report obtained from Experian Scorex PLUS; Prosper does not disclose the exact credit score of a borrower, but rather reports the so called Prosper Score or Loan Grade, which ranges from AA (Low Risk) to HR (High Risk) and is dependent on Prosper rating, expected loss rate, loan term, economic environment and competitive environment (Prosper – Investing, 2012). Apart from the Loan Grade, lenders can see historical delinquencies of a borrower, revolving credit balance, bank card utilization, the number of current credit lines, public records and credit inquiries.

The information that is voluntarily provided by the borrower, but could not be verified, is classified as 'soft' information. Examples of the latter include house ownership, employment status, expenditure flows, description of the loan purpose, personal or family pictures, etc. In addition to hard and soft information, Prosper facilitates leveraging on social information about the borrower, i.e. whether he/she is a member of a particular group on the platform, whether there are any friend endorsements, the width and deepness of the borrower's social network on the platform.

2.2 Offline Competition

Dhand et al. (2008), Greiner & Wang (2007; 2009) and Ryan et al. (2007) take a relatively extreme standpoint suggesting that peer-to-peer lending represents a disruptive innovation on the traditional lending market. Contrary to traditional lending institutions, a major attribute of all P2P lending platforms is the provision of hard information complemented with soft and social information in order to disclose a detailed profile of a borrower to potential investors. To a great extent researchers tend to agree that this feature represents a necessary prerequisite for peer-to-peer competitiveness relatively to traditional banks, particularly due to additional information leading to lower interest rates and expansion of credit markets (Wu & Xu, 2011; Ashta & Assadi, 2009; Herzenstein et al., 2008).

On a more conservative note, Heng et al. (2007) report the business perspective on the topic suggesting that P2P lending is likely to fill only a niche in the market of low rating loans. Yet, although currently traditional banking institutes and peer-to-peer lending platforms might be seen as complements to each other with respect to their target clientele, both of these lending forms will inevitably turn to direct competition. The research conducted by Meyer (2007) reconciles the opposite standpoints of Heng et al. (2007) and Dhand et al. (2008), and states that P2P lending platforms primarily target the segment of low rating borrowers that were previously excluded from the credit market, while banks are likely to move into the segment due to its high profitability. Effectively, although both banks and individual lenders prefer borrowers with a good rating with low risk profile, the only really profitable source in order to make money could be low rated borrowers. This profitability issue, therefore, leads to the direct competition between banks and P2P platforms within the low-rated borrower niche. However, Freedman & Jin (2011) seem to think exactly the opposite, i.e. that the P2P lending market and traditional credit institutes will approach each other in the high quality borrower segment; in their paper the authors conclude that lenders exhibit a steep learning curve which will inevitably rule bad borrowers out of the market. Even new lenders, who did not learn yet, will rely on the market's experience, consequently raising the average quality of funded loan applications. These findings also go along with those of Klafft (2008), who pointed out that borrowers with weak credit rating, who cannot get funding in the traditional banking system, will unlikely get any on the P2P lending market. Such a divergence in opinions over peer-to-peer future positioning could be explained by the relative novelty of the concept as well as by the rapidly changing market conditions caused by the recent financial crisis.

2.3 Relative Risks and Benefits of the Concept

2.3.1 Risks

The topic on which the majority of researchers tend to agree is that lenders represent the party that bears the risk linked with peer-to-peer lending transactions. The high risk perceived is the main reason behind the current undersupply of capital on online P2P marketplaces, and it represents the key issue for platforms to solve so as to ensure the future success of online social lending (Galloway, 2009; Rumiany, 2007).

Many are the risks that lenders bear when lending through P2P platforms. First and foremost, lenders replacing the traditional role of banks are exposed to the credit risk associated with the borrower, i.e. there is a potential for financial losses resulting from the failure of a borrower to perform their due obligations. Besides, in contrast with traditional savings accounts with fixed rates of return, loans on social lending platforms are not backed by any collateral or guaranteed by any third party: if, on one hand, platforms promote this features as an incentive for people to borrow via P2P marketplaces, on the other hand, this corresponds to the higher risk for lenders. As a consequence, in case of default, the return for a lender depends exclusively on the success of P2P lending platforms collecting agents in obtaining repayments from borrowers. Consequently, it further stresses the importance of choosing the right borrower when making an investment on the social lending market.

Moreover, since social lending platforms delegate payment collection to external agents instead of pursuing it by themselves, the operational risk steps in. The potential for unexpected losses due to inadequate or failed internal processes of P2P lending marketplaces is further increased by the fact that the information supplied by borrowers is not verified. For example, Prosper reported that it selected approximately 39% of loans listed from 2009 to 2010 for income and/or employment verification (United States Government Accountability Office, 2011). Since the borrower-related information might not accurately reflect his/her creditworthiness, the actual loan return that lenders get might be quite different from the promised one. As such, the mitigation of the operational risk represents a crucial challenge for P2P platforms, requiring significant reduction of information asymmetries between borrowers and lenders at the time of loan origination.

Apart from two major risk categories – credit and operational – lenders are exposed to market, liquidity and legal risks. Due to the absence of fines for early repayments, reductions in prevailing interest rates on alternative funding markets could induce borrowers to repay their loans in advance, thus affecting lenders' return. The upward trend on alternative markets still negatively affects investors return due to increased opportunity cost of an investment in peer-to-peer loans. This could partly be attributed to liquidity risks because on most of the social lending platforms there are no secondary markets for transferring loan ownership. Even if the secondary market option is provided, it is rare that the liquidity level is sufficient for efficient market functioning. Finally, the first peer-to-peer lending platform was founded in 2005, which makes the concept quite young and subject to potential regulatory changes as it gains weight. The latter could also undermine all benefits of peer-to-peer lending eventually decreasing its attractiveness over traditional banking institutions.

Since peer-to-peer lending in a nutshell is a financial marketplace, the aforementioned risks are based on same key problems as the traditional banking system – adverse selection and asymmetric information. Because lenders and borrowers remain anonymous to each other during the transaction on peer-to-peer platforms, the magnitude of these problems is larger in social lending, which presents itself through both ex ante adverse selection and ex post moral hazard (Akerlof, 1970; Pauly, 1974). The studies by Herzenstein et al. (2008), Freedman & Jin (2011), Lin et al. (2009), Berger & Gleisner (2009), Iyer et al. (2011) specifically show social lending vulnerability with respect to adverse selection and asymmetric information. With higher risk of participation in the social lending transaction the lenders adjust the expected rate of return upwards. This, eventually, squeezes healthy borrowers out of P2P lending market and forces those who are left to report false information that could not be verified. Nevertheless, there is a number of actions that peer-to-peer lending platforms undertake to minimize the effect of adverse selection and asymmetric information. For mitigating the effect of the adverse selection, to some extent P2P platforms cut off the tail of loan applicants with lowest credit scores by establishing a floor for the latter; and to cope with asymmetric information, P2P platforms provide investors with extra sources of information on the borrower, which includes hard and soft factors as well as loan auction details.

Consequently, Freedman & Jin (2011) and Iyer et al. (2011) among others wondered if the proposed solutions to both fundamental problems of social lending markets have any effect. Both papers prove that based on provided extra information, lenders can infer the true creditworthiness of loan applicants. In particular, Freedman & Jin (2011) find that investors impose up to 0.4% interest rate premium from 2007 onwards within the credit score group assigned by a P2P platform for those borrowers that just reached the better credit rating. Iyer et al. (2011) at the same time discovered that about 33% of the real credit rating difference within a category is inferred by the lender. The authors suggest that the verified and directly observable financial information play the biggest role in correct determination of the true credit score; however, they do not exclude the subjective part of decision making. The influence of the subjective part is found to increase with decreasing credit score. For example, Freedman & Jin (2011) as well as Herrero-Lopez (2009) found empirical evidence of the positive effect of social networks, and group affiliation on minimizing information asymmetries on peer-to-peer platforms. Moreover, in absence of the traditional intermediary that was generating the information about borrower creditworthiness, the transparency of loan auction and resulting herding behavior of investors are also found to be a significant source of information about the loan application. Herzenstein et al. (2010) defined herding behavior in online P2P lending as a greater likelihood of bidding in auctions with more existing bids, and although the researchers agree on the fact that investors extract additional information from the bidding process on the loan application, the effect of this behavior is dubious. Ceyhan et al. (2011) found that herding is negatively correlated with lender profits, while Herzenstein et al. (2010) claim herding to be beneficial as only loans which received full funding come into existence. The detailed description of the effect of verifiable, non-verifiable and loan auction data on loan fundability and default rates will be discussed in the following section.

2.3.2 Benefits

In his paper Slavin (2007) summarizes the benefits of peer-to-peer lending as substantially lower overhead and administration charges, zero interest expense and minimal regulatory costs. Apparently, the opportunity to avoid banks, reputed costly middlemen, provides the greatest support for social lending fans. Despite peer-to-peer lending platforms survive by charging servicing fees for every funded loan, fees do not exceed 2% - making the P2P option cheaper than traditional financial institutions. Apart from lower transaction costs, both parties benefit from increased transparency of the costs associated with borrowing/lending and the control that direct borrowing/lending gives them. Furthermore, Slavin (2007) suggests additional mutual benefits of peer-to peer lending stemming from the fact that people deal with real individuals and not faceless institutions while still maintaining anonymity of both parties.

As previously mentioned, the natural result for a borrower from avoiding the costly and nontransparent middlemen is the decrease in the borrowing rate. Being cautious not to scare away healthy borrowers, at this moment none of the peer-to-peer lending platforms imposes any early repayment punishments. This makes social lending particularly appealing to self-employed people and others with fluctuating earnings, which would negatively affect chances of getting a loan in a conventional bank. Besides expanding the range of borrowers eligible for funding, social lending platforms also increases the speed of the process for borrowers.

As online P2P lending marketplaces have leaner organization than traditional financial institutions, a typical P2P loan application takes three to four days to get approved, which is much faster compared to banks and other financing companies (Slavin, 2007).

Taking the lender's perspective, since skipping the banks as middlemen shifted borrowing rate downwards, it also shifted the return of lenders upwards. According to the study of Slavin (2007), the average lender on Zopa, a pioneer P2P lending marketplace based in UK, earns from 7% to 10% after bad debts are written off, which is twice what top British savings accounts pay. It also offers lender, in quality of investors, a high flexibility in the choice of target investment with respect to the interest rate, loan length and borrower characteristics. This flexibility eventually represents a good diversification opportunity for every investor, especially by providing a new asset class which is different from traditional bank deposits and stock market investments.

Summarizing the existing literature on benefits of the peer-to-peer lending concept for two key stakeholders, the key success for borrowers and lenders is determined by ex-ante and ex-post loan performances respectively. In other words, borrowers primarily benefit from higher ability to obtain credit, i.e. higher funding probability, and from lower interest rates; lenders, on the other hand, profit from better repayments of the loan, i.e. lower default probabilities, and from sufficient compensation for the risk undertaken. The following sub-sections will discuss in detail the factors influencing these 4 crucial parameters.

2.3.2.1 Borrowers

As mentioned above, the major benefit of the peer-to-peer lending concept compared to the conventional banking system is the access to extra information sources, particularly soft ones, about the borrower. Since this information could be classified into hard, i.e. verifiable, financial; soft, i.e. self-reported, non-verifiable, data; and social networks (group belonging) data, there is an ongoing academic debate whether peer-to-peer lending platforms can increase the preciseness in determining borrower creditworthiness based on soft & social information, which is usually not available for traditional banks. In the borrower's eyes, this translates into identifying all factors that can increase the likelihood of being funded and of being charged a lower interest rate.

Impact of borrower's hard, financial (verifiable) information

There seems to be a consensus among researchers regarding the importance of hard information on both the listing and the borrower characteristics (Ryan et al., 2007; Duarte et al., 2010). Herzenstein et al. (2008), for example, in their earlier work find that loan decision variables such as loan amount, maximum interest rate and duration of loan listing act as primary mediators between the borrower characteristics and the likelihood of funding success. The authors also suggest that overall financial healthiness indicated by loan grade, debt to income ratio and home ownership plays an essential role in funding decision-making. Ravina (2008), in her analysis of 11,957 loan requests, argues that investors rationally treat the provided financial information, preferring higher credit scores over any other financial or soft data when making the lending decision; analyzing over 17,000 Prosper funded loans, Miller (2011) also supports these findings. Klafft (2008), on the other hand, found that the existence of a borrower's bank account is a more powerful determinant of the loan funding success and of the final interest rate than the credit rating. Indeed, Klafft (2008) argues that investors do no use information rationally or lack financial expertise. Weiß et al. (2010) provide empirical evidence that investors on peer-to-peer platform tend to be risk-loving as they tend to fund more loans with higher expected losses and apparently preferring higher debt-to-income ratio. Barasinska & Schäfer (2010) join their conclusions explaining the surprising attractiveness of high debt-to-income ratios with investors' belief in the disciplining nature of high debt burdens when the portfolio of loans is highly diversified.

Impact of borrower's soft, self-reported (non-verifiable) information

Loan description (text length and content)

An ample stream of research has focused on the issue whether the soft factors can act as a key factor in order to determine the probability of loan funding. Once financial data is disclosed thanks to external credit agencies, borrowers legitimately question whether they should disclose additional information, and whether such supplementary data would lead to a higher probability of getting the loan funded. As it appears, borrowers with a more complicated background tend to use the option of explaining their story to a lender so as to enhance the likelihood of a rational, informed decision. Consequently, the effort put into the loan application, which could be proxied by the amount of extra non-verifiable information disclosed in the loan description (measured by length), should be theoretically correlated with a higher funding probability. Herzenstein et al. (2008) together with Meyer (2007) support this hypothesis. Yet, the paper of Böhme & Pötzsch (2010) does not fully agree with previous findings, as the authors find support for the hypothesis for all loan purposes except for private loans. This suggests that not only the length, but also –and primarily – the content of the loan description (i.e. the content of the additional information provided) acts as determinant of funding probability.

As the next step, other academics have focused on the description of personal characteristics and their potential effect on funding and interest rate charged. Herzenstein et al. (2011) in their later work made an extensive study analyzing 1,493 loan listings posted by borrowers on Prosper with the aim of trying to determine whether personal characteristics mentioned in the loan description have a considerable impact. The authors distinguished 5 identity claims commonly used in loan applications: being trustworthy, hardworking, successful, religious and affected by economic hardship (Herzenstein et al. 2011). Empirical findings show that the usage of several identity claims in the loan description has a significant positive effect on the probability of funding, while having no identity results in a negative effect. Everything else held constant, stating to be trustworthy or successful had notable positive effects, but stating to be religious had a negative one. Additionally, the authors found a correlation between worse credit ratings and the usage of identity claims, which supports their initial findings of the positive impact of the effort put into the loan application on the funding probability.

The final group of studies focusing on loan description content has focused on the impact of describing the loan purpose on the loan funding probability. Herzenstein et al. (2011) complemented this area of studies by revealing that two combinations of loan listing argumentation – explanation-acknowledgement and explanation-denial – increase the likelihood of favorable lending decisions. In this direction, Pope & Sydnor (2011) provide evidence that loans aimed at debt consolidation, i.e. if borrower states that he/she wants to pay down credit card debt with the loan from peer-to-peer lending platforms, is preferred over other loan purposes. This is said to be primarily due to attractiveness of interest rates on peer-to-peer lending platforms relatively to the ones for payday loans in traditional banks. Nonetheless, other authors provide evidence for contrary conclusions. For example, Caldieraro et al. (2011) found evidence of counter-signaling theory on peer-to-peer lending markets. After analyzing 26,335 loan applications, they concluded that a significant non-monotonic relationship exists between the non-verifiable (soft) information voluntarily provided by the borrower and the loan approval: more specifically, when borrowers withhold non-verifiable information, they are more likely to have their loan applications funded.

Appearance (race, gender and beauty)

Unlike the heavily regulated traditional banking system, P2P lending platforms, with their animosity, allow for considering available personal characteristics as the basis for discrimination of certain borrowers. Needless to say, this area attracted a notable number of researchers aiming at analyzing social lending markets on the bias towards funding loans of applicants with certain characteristics.

The primary factor for discrimination on peer-to-peer lending platforms is the race of a borrower (Herzenstein et al., 2008; Pope & Sydnor, 2011; Ravina, 2008). Pope & Sydnor (2011) show that the chances of loan listings of Afro-Americans to get their loan requests fully funded are 25 to 34 percent smaller than those of white skinned applicants, keeping the borrower credit rating constant. These findings go along with the work of Herzenstein et al. (2008) and Ravina (2008), who also provide empirical evidence that Afro-Americans have lower funding probability and are on average charged higher interest rates than representatives of any other race. In her later work, Ravina (2008) tries to address the reasons behind discrimination instead of the fact of its existence. Her results show that social resemblance has a strong positive impact on the lenders' decision. Living in the same city as the lender, belonging to the same ethnicity, gender, interest group or hometown all increase the funding probability and decrease the final interest rate.

The second major source of discrimination in social lending is the gender of a potential borrower (Ravina, 2008; Pope & Sydnor, 2011; Duarte et al., 2010). According to studies performed on the American P2P platform Prosper, women are more likely to get their loans funded on P2P platforms than men.

Barasinska & Schäfer (2010), however, present contradicting findings from the German social lending platform Smava: the main conclusion of their paper implies that, at least in Germany, gender has no effect either on funding probability, or on the final interest rate charged. This is contrary to works by Herzenstein et al. (2008) and Pope & Sydnor (2011), but in line with Böhme & Pötzsch (2010), who also analyzed a Smava data set. Thus, one cannot generalize that, everything else held constant, gender discrimination is a common feature of P2P lending markets.

Finally, unless personal information is disclosed directly by the borrower in the loan application, the only basis for discriminatory actions from the lender side is non-verified pictures. Along this path, Ravina (2008) touches upon more subjective grounds for discrimination, among which the 'beauty' of a potential borrower is the most interesting. The author found that beautiful borrowers are both 1.41% more likely to get a loan funded and pay 0.81% lower interest rate than an average-looking borrower, holding other factors constant. Yet, borrowers that are overweight, just as those that appear creditworthy, are more likely to get a loan (but do not pay lower interest rates). The findings of Ravina (2008) do not support the research outcome of Klafft (2008) or Pope & Sydnor (2011), who argued that investors are less likely to fund a loan and to demand lower interest rate from a borrower that appears rather unhappy or overweight. Pope & Sydnor (2011) also argue that age is an additional discrimination parameter extracted from borrowers' profile picture. Compared to a base group of 35-60 year-old loan applicants, there is a 0.4% to 0.9% higher chance of getting the loan funded for those who appear to be younger than 35. The findings of Böhme & Pötzsch (2010) support as well the hypothesis of age discrimination by providing evidence that older people on average pay higher interest rates.

Impact of information on social networks and group ratings

The final type of information that, unlike traditional banks, peer-to-peer lending platforms offer borrowers to disclose is their, so called, social capital by either forming groups or establishing friend connections. Due to the recent emergence of online social networking websites, this aspect of social lending got the biggest attention of researchers, trying to gather evidence on how technology can help quantifying soft information and transforming it into credible signals of the borrower quality (Lin et al., 2009).

No matter how the social capital is proxied on P2P lending platforms, either through group affiliation or several layers of friendship networks, Lin et al. (2009) together with Greiner & Wang (2009), Herrero-Lopez (2009) and Freedman & Jin (2011) suggest that borrowers with the social capital are more likely to get their loan funded at lower interest rates. The rationale behind the influence of the social capital on funding probability is based on social stigma of default, which is defined as the disutility suffered by borrowers when friends learn about their default (Crocker et al., 1998). If social stigma costs matter, borrowers who perceive themselves as being likely to default should avoid forming friendships. Besides, these members should also not express a wish to join or not be accepted by groups on social lending platforms. When it comes to the effect of having friends on P2P lending platforms, Lin et al. (2009) and Krumme & Herrero (2009) provide evidence of a positive effect of the number of friends as well as active bidding on the friend's loan on the funding probability. As for group membership, Weiß et al. (2010), Krumme & Herrero (2009), Hildebrand et al. (2010) and Lin et al. (2009) suggest a similar positive relationship with the probability of obtaining a loan and of a lower interest rate. According to Herrero-Lopez (2009) and Ryan et al. (2007), affiliation with groups increases borrower's chances of being funded by a factor of 2 and 3 respectively; and according to the studies of Berger & Gleisner (2009), Greiner & Wang (2009) and Hildebrand et al. (2010), social affiliation significantly reduces the final interest rate. Qiu et al. (2010, 2011) empirically support previous findings based on US peer-to-peer platforms by discovering even more significant findings on the Chinese online P2P market PPDai.

As for the group rating, which is determined by a peer-to-peer platform based on the group loan performance in the past, Berger & Gleisner (2009) and Collier & Hampshire (2010) find a positive effect on funding probability but a little or no effect on interest rates. The authors also discovered a link between the size of groups and the average interest rates charged to its members: as a matter of fact, larger groups have a better peer-review; therefore, members of larger groups benefit from lower interest rates on their loans. These findings, however, are contradicted by Freedman & Jin (2011), who claimed that peer-to-peer platforms' incentives for group leaders have the consequence of artificially increasing the group size. Freedman & Jin (2011) together with Greiner & Wang (2009) thus think that the ratio of lenders to borrowers within a group has a much greater effect on interest rate than the group size, i.e. the greater the proportion of lenders, the lower is the interest rate for borrowers.

To the group belonging factor, a controversial importance in several researches is dedicated to group leaders, whose endorsement is found to have the biggest positive effect on both funding success and total number of bids (Ryan et al., 2007). Even if the group leader endorsement can increase the likelihood of funding (Kumar, 2007), it does not have any effect on the interest rate of the loan (Berger & Gleisner 2009). As shown by Berger & Gleisner (2009) and Collier & Hampshire (2010), only the active bidding by the group leader can decrease the final interest rate. Freedman & Jin (2011), on the other hand, claim that since group leaders receive a fee for each successfully funded loan, the bid by a leader should be perceived as a bad signal. Nevertheless, as in 2007 Prosper abolished these incentives for group leaders, the negative effect of group leader bids has not been spotted anymore.

2.3.2.2 Lenders

A considerably smaller number of papers treat the issue whether factors benefiting borrowers through higher funding probability and lower interest rates also result in benefits for lenders via lower default probability.

According to Ravina (2008), all financial variables have expected signs: better credit ratings, lower final interest rates, a higher salary and a lower debt-to-income ratio decrease the probability of default. Iyer et al. (2011) also shared these findings proving that the loan score assigned by peer-to-peer lending platforms is both highly related to underlying creditworthiness of a borrower and predicts the likelihood of default. Miller (2011) tried to look at the difference between profiles of borrowers who default early and borrowers who default later. His findings suggest that having open debt lines lead to a lower default probability, as in case of difficulties more financial resources are at borrower's disposal. Contrary to other studies, he finds that home ownership is associated with a higher default probability, which is not in line what investors usually consider when giving home owners more preference in funding a loan (Herzenstein et al., 2008).

The misalignment of expected and realized loan performances based on factors that lenders take into account before funding a request only increases if these factors are of soft and non-verifiable nature. In contrast to a positive effect on the funding probability, the effort and content that borrowers put into the loan application does not pay off for investors. Stating to be trustworthy or successful is a particular example of this phenomenon, when borrowers are expected to behave in the way they describe themselves while they do not (Herzenstein et al., 2010). The authors also find that being moral leads to a lower default probability, but being affected by economic hardship – to a higher one. Meyer (2007) supports the hypothesis that effort put in the loan request is false as the length of the application is negatively correlated with delinquency risks. Caldieraro et al. (2011) results go along with the latter findings, and they remain consistent with the counter-signaling theory on peer-to-peer lending markets. According to Caldieraro et al. (2011), borrowers that do not need to explain anything in the application description are also more likely to be creditworthy. Only the findings of Iyer et al. (2011) show contrary evidence to aforementioned papers by stating that the more extra personal information the borrower discloses by writing lengthy descriptions, the lower is the default probability of his/her loans.

Discrimination fans on social lending platforms were disappointed by the findings of Ravina (2008), as she proved age, gender and all other personal characteristics except beauty to have no effect on loan delinquencies. Beauty, on the other hand, lead to a default probability 3 times higher than the average. She also found that retired, part-time employed or unemployed loan applicants possess higher delinquency risks. Pope & Sydnor (2011) complement these findings with the evidence of military and Afro-Americans having higher default probability.

Nevertheless, previous research prove that social capital leads not only to higher trustworthiness in the eyes of investors, but also keeps the motivation for making loan repayment (Greiner & Wang, 2009; Lin et al., 2009; Ryan et al., 2007; Hildebrand et al., 2010; Krumme & Herrero, 2009; Klein, 2008). The conclusion of Freedman & Jin (2011) is that borrower's friends are better equipped to identify risks and trustworthiness because of the additional information they possess due to the personal relationship. Moreover, they argue that monitoring within member groups also provide extra incentive to timely repay loans.

Everett (2010) decided to take a step forward and test whether there is any difference between group influence on loan performance. His findings suggest that members of groups with personal relationships have lower default rates, which reflect the financial intermediation role of the group on a peer-to-peer lending platform. If a group member lends to another, who is either a friend, co-worker, neighbor or former classmate, this can be seen as a positive signal of the borrower's creditworthiness. In a way, these findings match Ravina's (2008) results on the recent shift of investor preferences towards borrowers with similar profiles (social resemblance). Hildebrand et al. (2010) took another perspective on the role of a group on a social lending platform. The researchers looked at the effect of the incentive scheme for a group leader in order to facilitate the loan origination on the loan performance. The authors found that groups where group leaders were charging a fee for helping borrowers to originate a loan were characterized with higher default rates. In other words, group leaders did not bother about the borrower creditworthiness, but they were rather placing the costs on other lenders. However, if the group leader committed herself by bidding to a group member loan, the latter is shown to outperform relatively to the rest of the applications. In contrast to Hildebrand et al. (2010) results, there are several studies that did not find any effect of bids by a group leader on the loan repayment (Berger & Gleisner, 2009; Klein, 2008; Kumar, 2007). This indicates that the positive signaling effect of their bidding and the induced lowering of the interest rates by the lenders might not be appropriate. Besides, Miller (2011) and Everett (2010) challenge the majority of studies by presenting the evidence that mere group membership leads to a higher default probability.

Considering the mismatch between ex-ante and ex-post influence of aforementioned factors on loan performance, the question whether investors initially ask for sufficient compensation for the undertaken risks becomes attractive. However, only some authors tried to explore returns of lenders on social lending platforms, and currently their findings are still heterogeneous. Kumar (2007) empirically proved that lenders behave rationally and charge appropriate risk premiums relatively to expected default rates. Herzenstein et al. (2008) share his opinion stressing the potential for a higher return on P2P lending platforms than in traditional investment instruments. Krumme & Herrero (2009), however, suggested that the maximum returns for investors are achievable only in the segment of loans with high credit rating. The more recent researches by Klafft (2008), Freedman & Jin (2011) and Ceyhan et al. (2011) introduced and supported the hypothesis of both relative unattractiveness of peer-to-peer lending market for the lenders and their irrationality. For example, according to Ceyhan et al. (2011), 21% of all lenders received a positive payoff while 54% lose money. At the same time, Klafft (2008) suggested 3 rules for P2P lenders in order to achieve positive returns – choosing borrowers that did not default previously, that have reported income at least 5 times higher than their interest payments, and that did not inquire for extra credit during last half year. If lenders would follow these 3 rules, their realized returns would have been higher than Treasury yields for all credit ratings except for HR.

2.4 Contribution of this Paper to Existing Literature

Concluding this section, several authors have focused on the relatively new phenomenon of P2P lending market and its peculiar characteristics. However, research in the field shows little homogeneity and is still at its early stages, characterized by a dominant presence of studies stemming from a psychological-behavioral background rather than from a purely financial one. In particular, many papers seem to have focused on discovering the factors impacting the loan funding likelihood and interest rate setting, reporting then, ex-post, whether such factors led to a positive loan repayment or not.

As noted in the previous section, very few researchers like Klafft (2008), Freedman & Jin (2011) and Ceyhan et al. (2011) only ascertain inefficiency in the current investment strategies of social lenders. Klafft (2008) made a first step towards the direction of exploring possible advancements in the investment strategies, which would increase the peer-to-peer investment attractiveness over other investment instruments. However, the author does not seek for the most efficient strategy, which would lead to the highest return on the peer-to-peer loan market. A step forward was made by Singh et al. (2009) with the attempt to construct the efficiency frontier of peer-to-peer loan groups formed by similar mean and variance. However, the research focused only on discount cash flow model of realized return calculation and hard financial factors as primary determinants of the latter.

This paper is thus intended to enrich the poor set of existing literature to date on the P2P market. The research is going to focus on the analysis of investor perspective of the peer-to-peer lending market in order to balance out the current bias in researcher attention towards borrower side of social lending. Furthermore, the hypothesis stated by Herzenstein et al. (2008) regarding potential superiority of P2P lending market over traditional investment instruments remains overlooked, which identifies the specific research gap that this paper is going to fill. By constructing the efficiency frontier for a peer-to-peer lending market and determining the most efficient investment strategy, the relative attractiveness of the latter would be compared with the stock market investments with respect to the return each of the alternatives can yield, adjusted for the risk undertaken. Thus, this thesis will aim at understanding whether the peer-to-peer lending market could be considered as a profitable investment instrument, what should be the expected return for a rational lender and whether the risk-return ratio for an investor is high enough, compared to the equity market investment, to justify any active lending in the P2P platforms.

This work also aims to improve the efficiency frontier model previously built by Singh et al. (2009), complementing the peer-to-peer loan efficiency frontier modeling with soft information, which was proven to have equivalent importance on P2P markets to hard information, with overlooked individually forecasted default probabilities for each loan, charged-off loans and historical recovery rates. The usage of an extended data set with a timeframe ranging from 2007 to 2012 should also be considered as additional value of this research. This extension will allow a more careful evaluation of peer-to-peer loan investment attractiveness from the lender perspective.

It is worth noting that the sole focus on lender perspective, leaving aside borrower funding success, is not the only research scope limitation of this work. Considering the outlined characteristics of peer-to-peer lending platforms, two other emerging finance-related markets on the Internet should also be mentioned –

crowdfunding and microfinance. Yet, despite their operational similarity, they are conceptually different from the scope of P2P lending, and for this reasons not further explored in this thesis. Crowdfunding is indeed primarily focused on charity (which explains the popularity of the term 'nonprofit crowdfunding') via sponsoring private and business projects for the social benefit; microfinance is uniquely based on provision of financial services to low-income inhabitants in order to help poor people get out of poverty (Srinivasan & Sriram, 2003). This research, however, focuses only on peer-to-peer lending platforms that are dominated by investors with profit maximizing motives, excluding platforms driven by charity motives – i.e. crowdfunding and microfinance platforms. Such exclusion is justified by a resulting higher external validity and greater contribution of the research findings in relation to the research literature on household and individual investment decisions.

3 Methodology

Based on the spotted gap in the existing literature, this paper intends to study the attractiveness of peer-to-peer lending market as an alternative investment platform to other financial instruments. Therefore, the research question to be answered can be formulates as follows:

Under which conditions could peer-to-peer lending platforms offer relatively higher risk remuneration than traditional risky investments if at all?

The previous research by Klafft (2008), Freedman and Jin (2011) and Ceyhan et al. (2011) already suggested that lenders on peer-to-peer lending markets act irrationally, which leads to loss-making investment decisions. A more natural way of addressing the research question by comparing currently realized returns by social lenders and traditional risky market investors would not be correct due to mentioned difference in rationality of these two types of investors. Consequently, this research will base the comparison of returns of the most traditional risky investment instrument – equity markets – and peer-to-peer loan markets, if P2P lenders were able to rationally and effectively assess all the risks associated with a P2P loan investment. Furthermore, the research takes the Sharpe Ratio as the basis for comparison of investment alternatives because the ratio has been initially designed for comparing the attractiveness of investment opportunities in terms of risk remuneration. According to Sharpe (1966), the ratio is calculated as follows:

$$Sharpe\ Ratio_i = \frac{E(r_i) - r_f}{\delta_i} \tag{1}$$

where r_i is the return of the investment opportunity i, δ_i – the risk associated with the investment and r_f – the corresponding risk-free rate. If on the stock market the risk is associated with price volatility, then on the peer-to-peer loan market the key risk is the default probability until the loan reaches its maturity. The risk-free rate is the return on the most liquid and default free asset in the corresponding country, which in this case is the return on 1-month Treasury Bill of United States of America. Following the similar logic, S&P 500 is chosen as the representative stock market for the comparison with the peer-to-peer market in the U.S. Both of the choices are implicitly suggested by Kenneth French (French Data Library, 2012).

The main hypothesis of the research stems from the findings of Klafft (2008), Freedman and Jin (2011) and Ceyhan et al. (2011) and aforementioned specifications:

The Sharpe Ratio of the peer-to-peer loan market is lower than the one of S&P 500 for the sample period assuming that investors strive for achieving the optimal return.

Considering the abovementioned, methodology designed to address the research question and to test the hypothesis is divided into 5 consequent stages.

Stage I is concerned with determination of the first unknown in the Sharpe Ratio – the risk associated with the investment into a P2P loan. The borrower default being the main risk, the first stage of the analysis intends to find determinants of a loan default rate.

Stage II tries to calculate the second unknown of the ratio – expected return that a lender should expect from a P2P loan investment based on borrower and loan specific characteristics, default and recovery rates, supposing that lenders are rational and efficient in interest rate setting.

By combining results of two previous stages, **Stage III** aims at finding out whether there is any possibility of peer-to-peer lending markets to offer higher risk remuneration in the most optimal portfolio allocation case with the widest loan selection over the whole sample period. Therefore, based on the result of loan classification into loan classes through regression tree analysis, the efficiency frontier is constructed and the most optimal Sharpe Ratio of P2P lending market is determined.

Stage IV relaxes the first assumption made in Stage III by breaking down the sample period into monthly data and adjusting Stage III to actual monthly P2P loan availability on the market. The performance of P2P lending and stock market for shorter investment horizons is then compared.

The last stage, **Stage V**, relaxes the second assumption behind Stage III results that investors are efficient in interest rate setting and checks the validity of Stage III and Stage IV results by using the interest rates initially set by lenders for assessment of the relative attractiveness of a P2P loan investment.

3.1 Stage I: Default Rate Determinants

Over the last two decades researchers have developed numerous models for the estimation of the default risk and its determinants. These can be separated into two main categories: the first set unites structural models that build on the option pricing model approach suggested by Merton; and the second set – reduced form models. The former considers company/individual cash-flow as the stock price and debt level as the exercise price. Despite the strong assumption of complete information, the former class of models is characterized by the strong link between economic theory and statistical processes, which, in turn, increases interpretability of the output (Arora et al., 2005). The contrary could be said about reduced form models – the second major category of models for default probability estimation. These models exhibit high flexibility, which is both their strength and weakness. Since reduced form models do not set restrictive assumptions for obtaining higher predictive fit, the output obtained from purely statistical operations would suffer from disconnectedness with economic theory (Arora et al., 2005). The debate on the empirical preciseness of default probability forecasted by structural or reduced form models is not yet resolved, while researchers continue to provide contradicting results for both model sets.

However, as noted by Andreeva et al. (2007), the Merton model is known to be superior for predicting default probabilities for a short time horizon up to 1 year. Since the minimum investment horizon on peer-to-peer lending platforms has been 3 years has been complemented with other loan maturities only recently, the reduced form approach would provide more accurate results. Besides, the access to the information on all loans and borrowers is highly restricted by privacy policies of peer-to-peer platforms, which hinders complete information assumption behind structural models. Cetin et al. (2004) also suggest that in imperfect information conditions the use of reduced form models is more appropriate.

Within the reduced form category, most of the models for failure prediction employed probit/logit techniques. The latter are designed to determine the probability of a bank, company or private falling into one of two possible states of nature – default or non-default. The Cox proportional hazard model takes this approach a step further as it allows predicting not only the probability of default, but also the estimated time of this event. Additionally, proportional hazard models do not require the assumption about the distributional properties of the data (Whalen, 1989). Recent researches by Shumway (2001) and Chava & Jarrow (2004) also provide empirical evidence of the Cox proportional hazard model superiority in default risk modeling.

$$h(t) = h_0(t)\exp(x_i^T, \beta) \tag{2}$$

$$\log\big(h(t)\big) = \alpha(t) + \beta_1 x_{1i} + \cdots + \beta_k x_{k,i} \tag{3}$$

Formula (3) is the log transformation of formula (2) – the base specification of Cox proportional hazard model. The baseline hazard function, $\alpha(t) = \log(h(t))$, is the same for all borrowers but varies over time. The model does not make assumptions about the shape of $\alpha(t)$, but it assumes a parametric form on the effects of observed characteristics. However, it assumes that the effects of borrower *i* specific factors *k* denoted by $x_{i,k}$ on the dependent variable – time till default – are separable from the temporal effect embodied in $\alpha(t)$. The model allows censoring observations for which default never happened. Since loans on a peer-to-peer lending platform exist with different maturities, the dependent variable would be expressed as the proportion of months until default to the full term of the loan. Therefore, observations with time till default equal to 1 will be censored. Regarding risk probability determinants, previous research has proven that on P2P lending markets soft information about the borrower is as good in predicting loan performance as hard financial information (Duarte et al., 2010). The factors initially available for Prosper include loan amount, term, grade, loan description, purpose dummies, borrower group rating, debt-to-income ratio, inquiries in last 6 months, revolving balance and its utilization.

Two additional variables are constructed based on the previous literature: length of the description (Herzenstein et al., 2008; Meyer, 2009; Iyer et al., 2011) and the maximum rate the borrower is ready to pay (Iyer et al., 2011) as extra soft information sources. The construction of the former is based on counting the number of characters in the loan description written by the borrower. As for the latter extra variable, this research makes a distinction between the maximum rate set by the borrower and the maximum rate that he/she could have been charged judging upon similar borrowers and loans. Since the maximum rate set by the borrower acts as the actual interest rate cap, the distance between maximum rate that was charged from similar borrowers and the actual rate signals the collaterality of the requested loan as the result of availability of other funding sources and, thus, borrower's creditworthiness better than the absolute value of the maximum borrower rate. The construction of the maximum rate that could have been charged from the borrower is based on 3 following steps:

1) Determination of variables affecting the actual interest rate choice by running an ordinary least squares regression of the final interest rate actually charged from the borrower on loan and borrower specific variables;
2) Grouping loans with similar borrower and loan characteristics by sorting the universe of loans into buckets constructed in the N-dimensional space based on quartiles of N-number of significant interest rate determinants;
3) Calculation of the maximum interest rate for each loan group, which represents the maximum rate that the borrower could have been charged judging upon loans and borrowers with similar characteristics.

The Cox proportionate hazard model will produce two main outputs – the baseline hazard function for every period before the default and β_k effects of aforementioned parameters. These two outcomes will also allow for construction of a default probability function for every loan at every payment date before maturity.

The expected results are in line with the existing literature that all three categories of variables – hard and soft information sources as well as decision variables – should act as significant determinants of the default probability (Duarte et al., 2010). For example, more sound financial conditions of the borrower and availability of alternative funding represented by lower debt to income ratio, higher loan grade, higher absolute value of revolving balance and lower rate of utilization should be associated with lower default rates (Ravina, 2008; Iyer et al., 2011; Miller, 2011).

3.2 Stage II: Expected Return Calculation

Since the output of Stage I allows modeling the default probabilities for every payment period of every single loan with respect to loan and borrower specific characteristics, the expected return for every loan can be obtained from the binary tree of peer-to-peer loan scheduled payments.

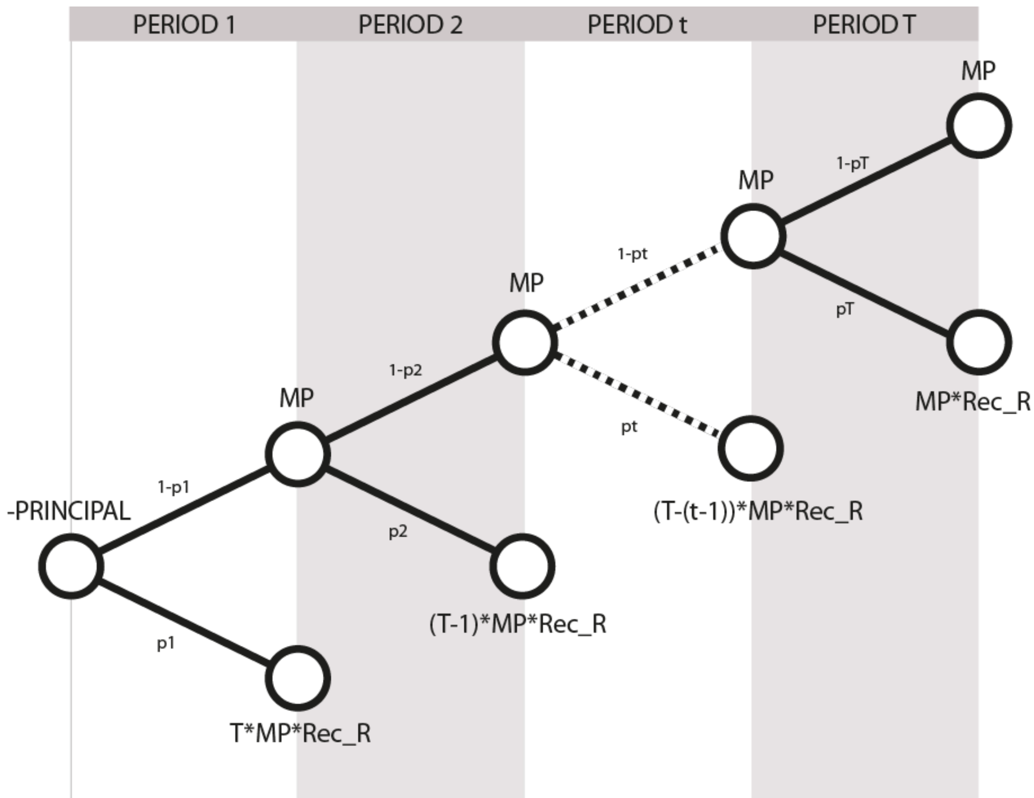

Figure 2 – Tree of Expected P2P Loan Repayments

The figure depicts the structure of repayments for every loan on Prosper platform with maturity T. Monthly payment (MP) is conditional to the probability of loan surviving by the specified period – $(1-p_t)$. Otherwise, there a p_t probability of default and loan repaying the recoverable part of remaining payments.

where $p_t$is the probability of default until time *t* (month #t since the loan has been issued), T – total length of loan term in months; Rec_R – average historical recovery rate for a peer-to-peer platform, $Principal$ – the total sum of money provided by lenders; MP – monthly payment, which on peer-to-peer lending markets is calculated as follows:

$$MP_i = \frac{(1 + r_i) * principal_i}{T_i} \tag{4}$$

where r_i stands for the final interest rate charged from the borrower.

The return for the risk that a P2P lender should get should be equal to the yield to maturity of the loan, i.e. the yield that makes the sum of discounted cash flows, adjusted to potential default probabilities in every repayment period as well as recovered cash flows by the end of the loan maturity, equal to initial principal provided by borrowers.

$$principal_i = \frac{(1-p_{1,i})*MP_i}{\left(1+\frac{YTM_i}{12}\right)^1} + \frac{p_{1,i}*T_i*MP_i*Rec\_R}{\left(1+\frac{YTM_i}{12}\right)^{T_i}} + \frac{(1-p_{2,i})*MP_i}{\left(1+\frac{YTM_i}{12}\right)^2} + \frac{p_{2,i}*(T_i-1)*MP_i*Rec\_R}{\left(1+\frac{YTM_i}{12}\right)^{T_i}} + \cdots + \frac{(1-p_{T,i})*MP_i}{(1+\frac{YTM_i}{12})^{T_i}} + \frac{p_{T,i}*MP_i*Rec\_R}{(1+\frac{YTM_i}{12})^{T_i}} \quad (5)$$

where YTM_i is the yield to maturity or the expected return of lender's investment into a specific loan i.

Following the suggestion by Klafft (2008), Freedman and Jin (2011) and Ceyhan et al. (2011) regarding the inefficiency of peer-to-peer lenders in terms of interest rate setting, if not specified otherwise, the expected return will be calculated based on the maximum interest rate for the similar loan and borrower group, determined in Stage I.

The results of this stage, in terms of breakdown of generated expected returns and Sharpe Ratios with respect to various loan characteristics, are expected to show that higher default probabilities are associated with higher required returns as well as to support the fact that average Sharpe Ratios of P2P loans are lower than that of the stock market, suggesting that on average peer-to-peer loan investments are unattractive comparing to the stock market with respect to risk remuneration.

3.3 Stage III: The Best Case Scenario on P2P Loan Market

Having generated the universe of varying expected returns in Stage II, Stage III tries to identify the most optimal investment strategy on the peer-to-peer market via efficiency frontier construction. However, the absence of a liquid secondary market for peer-to-peer loans or sufficiently long time series of expected returns of identical loans makes it impossible to construct the efficiency frontier in a traditional way. Instead, Singh et al. (2009) suggest first grouping existing loans with similar characteristics and distribution of returns and only then, considering these groups as loan classes and separate investment opportunities, constructing the efficiency frontier.

In order to assign every loan to a specific group with similar characteristics and expected return distribution, the classification and regression trees (CATR) approach is chosen (Singh et al., 2009). This conditional inference tee approach uses a unified framework of embedding recursive binary partitioning into well-defined theory of permutation tests developed by Strasser & Weber (1999) and (Singh et al., 2009). The CATR approach selects the split variable and split value at every node based on how good the association of dependent variable is with the independent variable. In this case, fair return of every loan acts as a dependent variable, and all observable to investor characteristics of the loan and the borrower – as independent variables. At the same time, the CATR does not require any assumptions (Timofeev, 2004) or subjective choice of the splitting variables in advance, as it determines only relevant ones from the provided set (Hothron et al., 2006). According to Timofeev (2004), the CATR is recommended for the analysis of financial data since it easily handles outliers by isolating them into a separate node without introducing bias to the overall result. However, drawbacks of the method include its ability to split the observations with respect to only one variable in each node and its sensitivity to adding/dropping variables or observations to the dataset (Timofeev, 2004).

As suggested by Singh et al. (2009), each group identified by the classification and regression trees approach will be perceived as a separate investment opportunity on the peer-to-peer lending market. Every opportunity will have its expected return, which is equal to the mean of the loan expected returns within a group, and its risk, which is equal to the group average of default probability up to the full loan repayment. Rationality should make investors require higher return for loans with higher risk; then, based on individual risk preferences, lenders can combine investment opportunities into portfolios, i.e. loans from different groups, to achieve maximum return per given level of risk. The expected return and the variance of the portfolio are calculated as:

$$E(r_p) = \sum_{i=1}^{n} w_i E(r_i) \tag{6}$$

$$Var(r_p) = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i\, w_j Cov(r_i, r_j) \tag{7}$$

where $E(r_p)$ and $Var(r_p)$ are the expected return and the variance of a portfolio p, which was formed by investing w_i share of the total portfolio in a loan group i, which has the expected return of $E(r_i)$, variance of $Var(r_i)$ and covariance $Cov(r_i, r_j)$ with a long group j, which is equal to zero for all loan classes following the suggestion of Singh et al. (2009), since loan groups are disjoint with respect to decision rules.
To get the maximum return for a given level of risk, the following optimization problem must be solved:

$$\max_{w_1, w_2, \ldots, w_n} E(r_p) \tag{8}$$

subject to following constraints:

$$\sum_{i=1}^{n} w_i \leq 1 \tag{9}$$

$$w_i \geq 0 \tag{10}$$

The first constraint ensures that the sum of all shares in the portfolio sum up to 100%, while the second constraint ensures that there is no short selling, as this option is not available on any peer-to-peer lending platform at the moment.

The outcome of Stage III will be the list of CATR splitting factors that social lenders could look at in order to sort loans into groups with similar return distributions, which would be helpful for making their investment strategies more efficient and profit maximizing. Based on the grouped universe of P2P loans, the efficiency frontier will be constructed featuring portfolios that maximize the expected return for a given level of risk, i.e. Sharpe Ratio. The latter will be then compared with the Sharpe Ratio for the S&P 500 index over the sample period in order to conclude whether investors following portfolio optimization strategies could achieve higher risk remuneration by investing in peer-to-peer loans on Prosper platform instead of in the US stock market. Results are again expected to go in line with the existing literature by Klafft (2008), Freedman and Jin (2011) and Ceyhan et al. (2011) stating that investors on peer-to-peer loan markets could not achieve sufficiently high returns even to break even, which would also serve as the answer for the research question of this paper.

It is worth noting, that the outcome of Stage III will be constructed based on the best case scenario for the Prosper marketplace, i.e. calculations will include the whole sample period assuming that all loans are available for diversification purposes in every time the investor makes the investment decision setting the maximum interest rate the borrower could be ready to pay. These assumptions of 1) maximum loan class diversity over the long investment period and 2) lender efficiency in interest rate setting are made in order to analyze whether peer-to-peer loan market can at all yield sufficiently high risk remuneration relatively to the stock market, if the former market is developed to its best observable state and investors are effective at setting correct interest rates.

3.4 Stage IV: Adjustment for Historical Monthly Loan Availability

Since Stage III outcome supplies the frontier of the most efficient portfolios with the most optimal return that is possible to achieve on peer-to-peer lending market in the best case scenario, Stage IV relaxes the first of the assumptions made in Stage III.

According the Modern Portfolio Theory, the wider is the selection of the assets on the market, the greater are the diversification benefits and, thus, the greater is Sharpe Ratio of the most optimal portfolio on the efficiency frontier. Since the P2P loan market and Prosper platform in particular are still in the growth phase, the selection of investment opportunities on the platform has not been constant over the sample period. Therefore, it is worth of testing whether the P2P loan market can offer better risk remuneration on shorter time periods and limited selection of loan classes.

Therefore, the methodology of Stage III will be applied to every month of the sample period, yielding monthly Sharpe Ratios of the most optimal P2P loan market portfolios, which would be compared with the latter ratios for S&P 500 index for the respective months.

The output of this stage should contain the list of monthly Sharpe Ratios for S&P 500 and for the most optimal portfolios on Prosper P2P loan marketplace. The results are expected to exhibit even worse risk remuneration of P2P loan investments relatively to the one of stock market due to limited opportunities of diversification particularly in the first months of Prosper operations.

3.5 Stage V: Adjustment for Lender Irrationality

The final stage of the research relaxes the second assumption made in the best case scenario modeling for the peer-to-peer lending market – inability of lenders to set the appropriate interest rates for borrowers. The latter assumption, based on results of Klafft (2008), Freedman and Jin (2011) and Ceyhan et al. (2011) studies, allowed for using maximum chargeable interest rates instead of actual ones in expected return calculation and efficiency frontier construction.

The stage applies the methodology of Stages II-IV using the actual interest rates that were originally set by the lenders and acts more like a robustness test of the results obtained in the preceding parts of the analysis. This output should comprise adjusted outputs of previous stages and is expected to show even worse relative risk remuneration of P2P loan market comparing to the result in Stage IV due to inefficiency of interest rate setting by lenders. Thus, this stage should be the final to test previously obtained results for validity in answering the main research question of this paper.

4 Data Analysis

4.1 Data Sources

To conduct the analysis described in the previous section, data from Prosper.com has been used as the primary source. However, the access to the full information about borrowers and loan performance is limited to registered users on the platform, i.e. U.S. registered residents. At the same time, Prosper makes some data available specifically for purposes aimed at facilitating academic research on peer-to-peer lending. Since the platform does not allow full direct access to the data for non-U.S. citizens, this thesis acknowledges the potential incompleteness of the provided data with respect to certain borrower specific variables.

After eliminating loans that are currently in funding or are still in process of repayment as of April 19, 2012, the final dataset consists of 11,752 loans originated from June 14, 2007, to February 12, 2012, i.e. from the first loan funding date to the last non-current loan status update.

Table 1 summarizes the variables that were initially available in the dataset. Besides the data on loan performance indicators such as the current loan status, loan start and end dates, principal remaining and total payments made, information is provided on the amount borrowed, term of the loan, loan grade assigned by the platform, borrower rate, loan purpose, loan description, number of inquiries to the credit agency regarding borrower credit status and history, revolving credit card balance, revolving credit card utilization, debt to income ratio and group rating. Based on the parameters initially available, two extra variables were constructed – loan description length and the difference between the maximum and actual rate paid by the borrower.

Data on historical dynamics of S&P 500 and the risk-free rate over the sample period, on the other hand, is obtained from the Kenneth French data library.

Table 1 – Variables and related description

Variable name	Description
	A. Peer-to-Peer Lending Market Variables
Amount	The monetary amount of borrowed money
Borrower Rate	The final rate set to be paid by the borrower
Debt-to-Income Ratio	Total debt to income ratio of the borrower at the time of loan origination
Group Ranking	The ranking, from the highest (1) to the lowest (12), of the group that the borrower is a member of on the P2P lending platform. The rating is assigned by the platform based on the repayment history of group members.
Inquiries Last 6 Months	The number of inquiries to the credit agency regarding borrower credit status and history during the last 6 months
Loan Description Length	The symbol length of the loan description written by the borrower to explain the reason behind the loan
Loan Grade	The grade from AA (1) to HR (7) assigned to the loan by P2P lending platform based on the credit score of a borrower
Loan Purpose	Purpose of the loan, clustered as follows: Debt Consolidation, House Improvement, Business, Credit Card Refinancing, Education, Car, House, Major Purchases, Medicine, Relocation, Vacation, Wedding, Renewable Energy, Other
Revolving Balance	The monetary amount of revolving credit balance
Revolving Balance Utilization	The percentage of available revolving credit that is utilized at the time the loan is originated
Term	The term of the loan contract in months
Time till Last Payment	The time until the borrower made the last payment due to default or loan maturity as a share of the total repayment period. The variable takes the value of 1 if the loan did not default at all, and 0 if the loan defaulted straight away without a single payment.
	B. Other Market Variables
Risk-Free Rate	The monthly return on 1-month Treasury Bill of United States of America
Market Risk Premium	The difference between the return of S&P 500 index and the risk-free rate on monthly basis

The table is divided into two parts. Part A presents the summary of variables initially available for the sample of P2P loans with corresponding descriptions. Part B presents the variables necessary for Sharpe Ratio construction for the P2P loan and stock markets.

4.2 Descriptive Statistics

First and foremost, the composition of the dataset according to the different variables identified above is presented. Table 2 gives an overview of the sample data. Judging upon initial descriptive statistics, most of the loans on the Prosper P2P lending platform were issued with 36-month maturity. The small number of 60-month loans is explained by a rather recent introduction of this loan option on the platform. The amount lent on average has been around 10,000 USD at the average borrower rate of 12.33%. The average borrower on Prosper platform had a Debt-to-Income ratio of 12.46% and a Revolving Balance of 15,088 USD, of which 46.67% were already utilized. Prior to applying for a social loan he/she already inquired for credit 1.52 times during last 6 months. He/she wrote a 344.5 symbol long loan description in order to get a P2P loan, 80% of which is usually paid out.

Table 2 – Descriptive statistics

Variable	Mean	Min/Max	Std.Dev.
A. Peer-to-Peer Lending Market Variables			
Amount	9,787	500/35,000	6,456.6
Borrower Rate	12.33%	5.42%/23.52%	3.22%
Debt-to-Income Ratio	12.46%	0/29.96%	6.88%
Group Ranking	6.633	3/12	1.658
Inquiries last 6 Months	1.5192	0/33	2.09
Loan Description Length	344.504	0/3,968	460.554
Loan Grade	5.116	1/7	1.485
Revolving Balance	15,088	0/1,207,359	29,449
Revolving Balance Utilization	46.67%	0%/119%	29.44%
Term	39.28	36/60	8.24
Time till Last Payment	0.798	0/1	0.340
B. Other Market Variables			
Risk-Free Rate	0. 0695%	0/ 0.42%	0.11%
Market Risk Premium	0. 21%	-18.55% /11.53%	5.96%

The table is divided into two parts. Part A presents the descriptive statistics of variables initially available for the sample of P2P loans. Part B presents the descriptive statistics for variables necessary for Sharpe Ratio construction for the P2P loan and stock markets. Mean column shows the average value, Min/Max – minimum and maximum values, Std.Dev – standard deviation of each variable.

Loan grades are not normally distributed, as it can be easily seen in Table 1 (Appendix A). They are skewed towards higher rating – 69% of all funded loans had AA, A or B rating. This suggests that lenders on Prosper were more risk-averse and preferred loans with higher ratings. Since loan ratings are based on credit scores obtained from the credit agency, the lower loan ratings are associated with higher risks. Thus, as Table 1 (Appendix A) suggests, the average interest rate, debt to income ratio (DTI), the number of inquiries during the last 6 months, the revolving credit balance and its utilization are relatively higher, while group ranking is relatively lower for borrowers with lower loan grades. The length of the loan description tends to increase with decreasing credit score of the borrower, who seems to utilize the opportunity provided by P2P lending platforms in order to explain his/her difficult financial situation to potential investors. Loan term and amount borrowed also show an upward tendency with lower loan grades, suggesting that healthy borrowers are less needy of capital and at the same time they are more certain and stable about their income for the foreseeable future.

Table 2 in Appendix A summarizes the dataset clustered by loan purpose. The majority of loans were issued for financial purposes – either for debt consolidation (42.92%) or for refinancing credit card debt (12.00%). This trend can be justified by the widely promoted opportunity to obtain a loan at a lower interest rate on P2P lending platform instead of relatively higher interest rates in banks; this positioning could have attracted the borrowers particularly with this loan purpose, which created such a disproportionate distribution. Concerning the loan purposes and loan ratings taken together, Table 3 in Appendix A suggests that borrowers with higher ratings are applying for P2P loans for enhancing their lifestyles – new house or car acquisition, house improvement, education, vacation, etc., more often than borrowers with lower loan ratings. On the other hand, lower loan ratings were mostly aimed at debt consolidation and starting small business.

Talking about unpaid loans (Table 4 in Appendix A), the latter constituted 29.15% of the sample and could be characterized by similar qualities as low rating loans. The average amount raised, the number of inquiries, group ranking revolving balance and its utilization of unpaid loans are relatively higher than the same parameters of fully paid loans. The lengthy loan descriptions do not seem to help in reducing interest rates, since Table 4 (Appendix A) shows that lenders anticipate the likely default and charge higher interest rates. As for the loan purposes (Table 5 in Appendix A), both paid and unpaid loans are still dominated by debt consolidation (42.27% and 44.48% of fully paid and unpaid loans respectively). The only two notable differences are with respect to loans aimed at starting small business and credit card refinancing. There is a twice larger share of business loans that were not paid in full relatively to those that were; this situation could be potentially created by the limited liability nature of newly registered enterprises. On a positive note, there is a 4.66% larger share of loans with credit card refinancing purpose that were paid in full relatively to those that the credit card loans were not. Such an observation could be explained by relatively lower interest rates available on P2P lending platforms relatively to usually high payday loan and overdraft rates in banks.

Consequently, considering the findings shown in Table 2 and Table 5 (both in Appendix A), one can conclude that, despite almost twice longer descriptions, almost every second loan funded with business purpose is not paid in full, in contrast to every fifth – with credit card refinancing purpose. Meanwhile, the lenders seem to anticipate this risk and charge on average the highest interest rate (13.48%) relatively to other loan purposes. Car loans are characterized with the longest terms, higher quality borrowers and, in turn, lower interest rates; this is also rewarded with better loan performance – car loans have the largest share of issued loans that were paid in full. Similar above average performance is demonstrated by loans for house or other major acquisitions. These findings could be explained by the tangible nature of these loan purposes, which in case of default would be relatively easier to convert back into capital than expenses for vacations or medical treatments.

To sum up, the preliminary overview of the available data suggests that investors have been rather risk averse on the Prosper P2P lending platform and have seemed to be able to spot categories with the highest proportion of unpaid loans and to charge higher interest rates for these loans. Meanwhile, borrowers mostly used P2P lending platform for debt consolidation and refinancing purposes. When it comes to loan performance, loans with tangible assets as purposes for borrowing seemed to have the highest proportion of fully paid loans, while loans aimed at starting up a small business – the lowest.

5 Empirical Results

This section presents results of applying the 5 stage methodology outlined in Section 3 on the empirical data from Prosper P2P lending platform described in Section 4. The presented results are structured according to the sequence of their gradual generation.

5.1 Stage I: Default Probability Determinants

Following the structure of the outlined methodology in Section 3 and utilizing the data described in Section 4, the first step concerns finding historical determinants of the interest rate in order to divide the loan universe into buckets with highly similar loan and borrower characteristics and to determine the maximum rates borrowers would have been ready to pay. For this purpose, an ordinary least squares (OLS) regression of the loan interest rate has been run on the available parameters: loan amount, term, grade, description length, purpose dummies, borrower group ranking, debt-to-income ratio, inquiries in last 6 months, revolving balance and its utilization.

Table 3 exhibits the regression output suggesting that 6 out of 10 aforementioned parameters resulted highly significant. Apart from the intercept, these are loan amount, term, grade, borrower group ranking, revolving balance and its utilization. In line with expectations, the indicators of higher riskiness of the loan are associated with higher interest rates (Lin et al., 2008). Similar sign has an effect of the group rating, which shows that poor performance of the group peers of the borrower serves as an indicator of his/her riskiness for investors that in the end demand higher interest rate compensation (Miller, 2011; Lin et al., 2009). Positive relationship between interest rates and the loan amount and the term could be explained by the term premium charged by investors for tying up more capital for the longer term, which is particularly important due to the absence of the secondary market for peer-to-peer loans. Both higher revolving balance and its lower utilization signal availability of alternative capital sources of the borrower, which he/she could employ in case of difficulties of meeting P2P loan obligations, thus, bringing the associated risk level and the interest rate down. The R^2 of the regression amounted to 97.31%, confirming the validity of the results.

Table 3 – Regression Outcome: Loan Interest Rate against Available Parameters

Variable	Coefficient	T-stat
Intercept	2.31 ***	17.00
Amount	0.0000065***	7.66
Term	0.01 ***	15.77
Description Length	0.000016	1.50
Loan Grade	-0.29 ***	-35.01
Inquiries Last 6 Months	-0.0029	-1.18
Revolving Balance Utilization	0.0017 ***	8.33
Revolving Balance	-0.00000057 ***	-3.23
Debt-to-Income Ratio	0.00032	0.41
Group Ranking	1.64 ***	215.87
Purpose: Debt Consolidation	0.095	0.94
Purpose: Credit Card Refinancing	0.036	0.35
Purpose: Business	0.105	1.02
Purpose: Education	0.076	0.73
Purpose: Car	0.034	0.33
Purpose: House Improvement	0.116	1.14
Purpose: House	0.151	1.39
Purpose: Medicine	0.063	0.59
Purpose: Major Purchases	0.091	0.89
Purpose: Relocation	0.123	1.14
Purpose: Vacation	0.117	1.03
Purpose: Wedding	0.135	1.28
Purpose: Other	0.092	0.91

The table presents the output of the ordinary least squares regression of the actual interest rate in percentages set for the loan on the borrower specific characteristics (Inquiries Last 6 Months, Revolving Balance, revolving Balance Utilization, Debt-to-Income Ratio and Group Ranking), loan specific characteristics (Amount, Term, Description Length, Loan Grade and Purpose dummies). The dummy for Renewable Energy loan purpose has been omitted. Coefficient column present the factor loadings, while T-stat values in the corresponding column pertain to a test of significance of the loadings. * - indicates 10% significance level; ** - indicates 5% significance level; *** - indicates 1% significance level. The R^2 indicator of the regression is 97.31%.

Having determined the key parameters of the interest rate setting, 2,048 unique loan bundles along 6 dimensions were created with highly similar loan and borrower characteristics. The maximum rate for the bundle will be used further in Stage II for the calculation of the maximum expected return. However, the difference between the maximum and the actual interest rate will also be used as one of the parameters in the default probability determination.

The higher value of the parameter is supposed to explain, from one side, the low level of necessity of the loan to the borrower and, from the other side, the lender inability to correctly set up the interest rate. Table 4 gives an insight into the distribution of the new variable across the initially given parameters. While there is an expectedly higher need for capital among people who ended up in default, a more interesting finding concerns higher parameter values for very high and low loan ratings. This might suggest that lenders tended to be overoptimistic about AA-loan borrower creditworthiness and to underestimate the risk associated with D-, E-, HR-loan borrowers, as implied in previous studies such as Herzenstein et al. (2008). Table 5 completes the picture by clustering the newly created variable by loan purpose.

Table 4 – Distribution of i(max) – i(actual) Variable across Fully Paid, Not Fully Paid Loans and Loan Grades

	Total	Fully Paid	Not Fully Paid	AA	A	B	C	D	E	HR
i(max) - i(actual)	1.466%	1.454%	1.493%	2.15%	0.69%	0.70%	1.38%	3.82%	2.72%	1.51%

The table presents the distribution of the difference between the maximum interest rate that could have been charged from a borrower judging upon similar loan and borrowers and the actual rate for the particular loan. Total column presents the average parameter value for the whole dataset, Fully Paid – for the subsample of loans paid in full, Not Fully Paid – for loans that either defaulted or were charged off during the repayment process. The rest of the columns correspond to the loan grade subsamples.

Table 5 – Distribution of i(max) – i(actual) Variable across Loan Purpose

	Debt Cons.	House Improv.	Business	Card Refin.	Education	Car	House	Major Purch.	Medicin	Relocation	Vacation	Wedding	Renew. Energy	Other
i(max)-i(actual)	1.50%	1.48%	1.45%	1.48%	1.35%	1.54%	1.33%	1.49%	1.51%	1.44%	1.52%	1.38%	1.13%	1.35%

The table presents the distribution of the difference between the maximum interest rate that could have been charged from a borrower judging upon similar loan and borrowers and the actual rate for the particular loan. Each column in the table corresponds to the loan purpose subsample.

As the second step towards modeling individual default probabilities, the difference between the maximum and the actual borrower rates, along with aforementioned borrower and loan specific variables, is used for construction of default probability default and survival base function via the Cox proportionate hazard model. All three categories of variables – hard and soft information as well as decision variables – found to be significant, supporting the findings of Duarte et al. (2010) that subjective variables just as hard financial factors have predictive power as determinants of default probability function. Table 6 summarizes the findings of Cox proportionate hazard model.

Table 6 – Cox Proportionate Hazard Model

Variable	Coefficient	Z-stat
Amount	0.0000038	1.35
Term	0.029 ***	15.36
Description Length	-0.000095 **	-2.50
Loan Grade	0.100 ***	3.25
Inquiries Last 6 Months	0.60 ***	8.96
Revolving Balance Utilization	0.006 ***	8.81
Revolving Balance	-0.00000052 ***	-0.93
Debt-to-Income Ratio	0.013	4.63
Group Ranking	0.251 ***	8.86
i(max) – i(actual)	-0.034 ***	-2.58
Purpose: Debt Consolidation	-0.650 **	-2.14
Purpose: Credit Card Refinancing	-0.923 ***	-3.00
Purpose: Business	0.018	0.06
Purpose: Education	-0.416	-1.29
Purpose: Car	-0.930 ***	-2.87
Purpose: House Improvement	-0.543 *	-1.76
Purpose: House	-0.673 **	-1.98
Purpose: Medicine	-0.384	-1.19
Purpose: Major Purchases	-0.703 **	-2.24
Purpose: Relocation	-0.432	-1.30
Purpose: Vacation	-0.263	-0.75
Purpose: Wedding	-0.815 **	-2.48
Purpose: Other	-0.423	-1.38

The table presents the output of the Cox proportionate hazard model with time until default as a dependent variable and the borrower specific characteristics (Inquiries Last 6 Months, Revolving Balance, revolving Balance Utilization, Debt-to-Income Ratio and Group Ranking) and loan specific characteristics (Amount, Term, Description Length, Loan Grade and Purpose dummies) as potential default probability determinants. The dummy for Renewable Energy loan purpose has been omitted. Coefficient column present the factor loadings, while Z-stat values in the corresponding column pertain to a test of significance of the loadings. * - indicates 10% significance level; ** - indicates 5% significance level; *** - indicates 1% significance level.

Considering the financial variables, all of the included variables turned out to be highly statistically significant except for the debt-to-income ratio. In the study of Miller (2011), a similar unexpected finding was observed with respect to the loan grade, which turned out to be insignificant in predicting future default probability of the loan. This is reasoned by the fact that loan grade is a complex factor, which is assigned by Prosper platform based on several financial variables. Therefore, as Miller (2011) had access to a much wider range of borrower specific financial variables, the loan grade effect was captured by the latter instead. In this case, since the exact mechanism of setting of the loan grade is not disclosed, the loan grade variable could have taken the effect of other financial variables including debt-to-income ratio. Besides, despite the correct positive sign of the relationship with the default probability, the debt-to-income ratio is a borrower self-reported parameter, the validity of which could be questioned in the nutshell.

Therefore, as Table 6 indicates, the signs of all financial variables except for the loan grade turned out to be in line with the existing literature. The higher number of inquiries for additional credit by the borrower during the last 6 months indicates the higher need for capital, which, in turn, results in higher default probability (Miller, 2011; Lin et al., 2009). Contrasting with the need for capital, revolving balance and revolving utilization exhibit the availability of alternative funding sources: the smaller revolving credit balance might signal unwillingness of the bank to provide more revolving credit to the borrower, implying a higher default probability (Miller, 2011). The higher the utilization of the revolving credit, the less untied resources are in borrower's disposition to meet debt obligations in case financial challenges reappear (Miller, 2011; Lin et al., 2009).

According to Table 6, the decision variables such as loan amount and term of the contract have expected positive relationship with the default probability. Although the loan amount is not a statistically significant component of the default probability base model, the sign of its effect is in line with the findings of Meyer (2007), Miller (2011) and Lin et al. (2009), suggesting that the larger is the actual value of the debt, the higher is the need for capital, the smaller is the probability that the borrower will be able to repay it. At the same time, the empirical evidence of positive effect of the loan term on the default probability is in line with Everett (2010): the potential explanation of these findings could be based on the fact that the longer is the time horizon, the higher is the probability of borrower specific and factor changes, which could eventually lead to a default. The final decision variable included in the regression is the difference between the maximum and the actual interest rate. The empirical evidence supports the suggestion by Iyer et al. (2011) that the maximum borrower rate is likely to serve as a credible signal of a borrower's creditworthiness. The statistical significance of the distance on which the borrower decided to set up the maximum rate relatively to the maximum paid by similar borrowers suggests that, indeed, this parameter is a good proxy for borrower creditworthiness, which is negatively related to the probability of default.

As to the highly debated soft information sources that in this research are represented primarily by the borrower group ranking and the loan description length, both of the latter parameters turned out to have a significantly positive effect on the survival probability of the loan. As suggested by Berger & Gleisner (2009) and Krumme & Herrero (2009), the group and particularly their leaders are considered as mini-intermediaries producing extra information about the borrower. In this case, the fact that the borrower is accepted to a high-ranking group consisting of members with successful repayment history sends a positive signal about his/her creditworthiness.

Among the bunch, the more debatable sign is the description length indicator. In contrast to Meyer (2007) finding that the more effort the borrower puts into the description of the loan purpose means that he/she needs to explain more, which eventually signals lower creditworthiness, the results of Cox proportionate hazard model in this research suggest the opposite. The more information the borrower includes into the loan description in order to decrease information asymmetries, the lengthier is the description and, thus, the more educated lender funding decisions should be. This should eventually mean lower default probabilities among funded loans with lengthy descriptions, which is supported by statistically significant empirical evidence presented by Iyer et al. (2011).

Finally, the results of this study identify 2 major categories within the possible purposes that turned to be statistically significant determinants of the default probability – debt refinancing and major purchases. Besides the obvious difference between loan rates available on the peer-to-peer lending platform and payday loans rates making the debt consolidation and refinancing a potential way to avoid default, the willingness to restructure the debt and meet the obligations with the other party might signal trustworthiness and high morale of the borrower and, in turn, his expected willingness to avoid default on the peer-to-peer lending platform as well. On the other hand, the group of purposes associated with major purchases such as car or house acquisition also push down the default probability, as the latter two present tangible assets that could be turned back into capital in case of difficulties to meet debt obligations.

5.2 Stage II: Expected Return Calculation

Having calculated the individual time-dependent default probabilities for every loan, this thesis proceeds with the expected return calculation considering the average historical rate of recovery. As this information is not available for Prosper, 18.2% recovery rate is taken from another major P2P platform, Lending Club, which operates on the same market as Prosper, has started operations with only a few-month difference and, thus, can be considered as a good proxy for the Prosper marketplace (LendingClub Loan Performance Page, 2012).

Overall, this study provides evidence of significantly positive expected returns of investment into peer-to-peer loan market based on Prosper platform. The average expected return from an investment is 24.43%. Table 7 exhibits the expected return breakdown across loan grades and purposes.

Being supported by the earlier study of Freedman & Jin (2008), these findings go in contrast to initially expected alignment for Klafft (2008), Freedman & Jin (2011) and Ceyhan et al. (2011) that investors lose money due to investments in P2P loans. According to this research, investors do not lose money per se, but rather make inefficient funding decisions since is the expected return disproportionate to the risk associated with the investments. Despite proving initially an expected positive relationship between the risk expressed as the probability of default and the return, the overall average Sharpe Ratio of 0.587 as well as the one for every loan grade category is still lower than the average Sharpe ratio of the stock market (1.25) during the same period of time. This evidence Moreover, judging upon results in Table 7, the Sharpe ratio is likely to be negatively associated with the risk and the alignment between the undertaken risk and the corresponding return is weaker for higher risk than for lower risk categories.

Table 7 – P2P Expected Return and related Sharpe Ratio

Loan Grades	N	Average Probability of Defaulting by end of the term	Expected Return	Sharpe Ratio
AA	2278	18.953%	15.664%***	0.757
A	3104	32.979%	21.951%***	0.626
B	2730	43.797%	25.369%***	0.549
C	1883	53.828%	28.726%***	0.509
D	1088	66.733%	33.200%***	0.478
E	441	88.301%	35.493%***	0.387
HR	228	96.361%	35.905%***	0.359

The table presents the distribution of the average probability of defaulting by maturity, expected return and the corresponding Sharpe Ratio across the loan grade subsamples, which consist of N-number of observations.

5.3 Stage III: The Best Case Scenario on P2P Loan Market

After using the classification and regression trees, 271 groups of loans with similar return characteristics were formed. Figure 3 depicts the higher end of the classification tree.

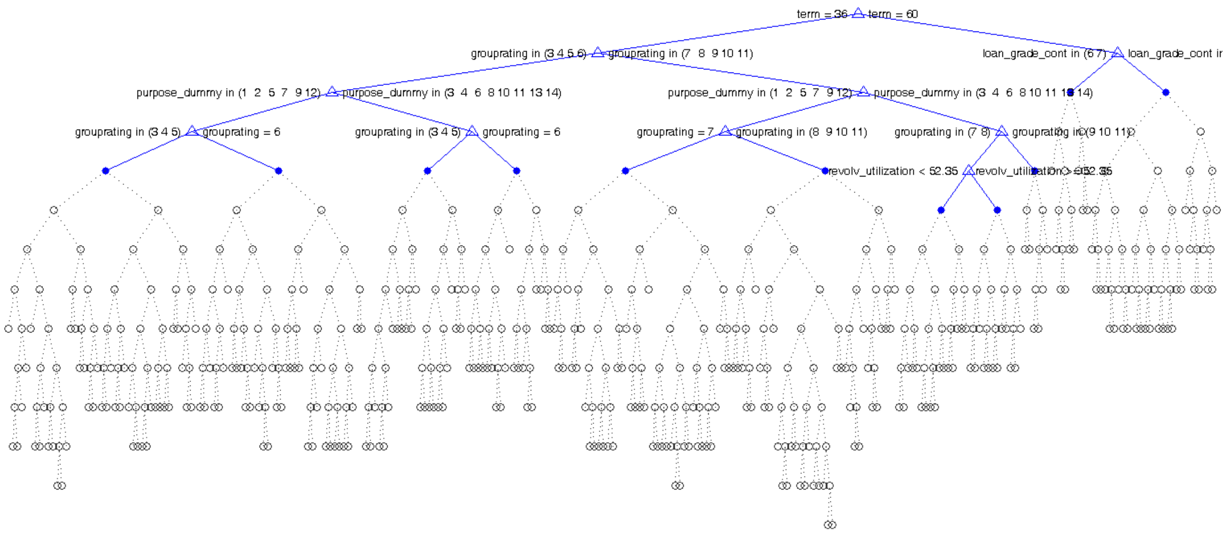

Figure 3 – Classification Tree

The figure depicts the output of the classification and regression tree analysis of the whole sample loan universe. The tree has been constructed based on group loans with similar characteristics of the main variable – expected return. For simplicity only the first splitting nodes and the respective decision criteria are presented. The total number of terminal nodes is 271.

Despite credit grade being the most used variable to analyze the loan performance on Prosper, results of this research are in line with the empirical evidence provided in the paper of Singh et al. (2009) that loan grade is not the primary variable to identify groups with similar risk and return profiles. In this case, the first classification variable is the loan term, which could be also explained by the fact that 5-year loan option has different risk-return characteristics from 3-year maturity loans. Besides, the long-term loan option has been introduced on Prosper only recently and, thus, is underrepresented in the sample. However, looking at the 3-year maturity loans as the basis of the analysis, the loan grade does not appear even among the 5 first splitting criteria. Instead, group ranking, loan purpose and revolving balance utilization are more useful in identifying loans with similar characteristics. These results again indirectly supports the hypothesis of Duarte et al. (2010) that soft information sources possess at least as much significance about the borrower on the peer-to-peer lending markets as financial data. Besides, it also suggests that lenders on Prosper should make their investment decisions not only based on loan grades, but also take other variables into account (Singh et al., 2009).

Based on the process of combining the loan groups, which represent investment classes on the P2P lending platform, into portfolios, the efficiency frontier for the whole P2P loan universe has been constructed. Figure 4 exhibits the constructed frontier. The tangency point of the capital allocation line between the risk-free rate, the average of which in the sample period equaled to 1.32%, and the efficiency frontier represents the most optimal portfolio with respect to the remuneration for the unit of risk. The most optimal return that could be achieved based on the results of this study is 42.64% with the undertaken risk level of 12.93%.

Consequently, it results in a Sharpe Ratio of 3.20, which is both higher than averages reported in Stage II and higher than the stock market Sharpe Ratio (1.25) for the same period of time.

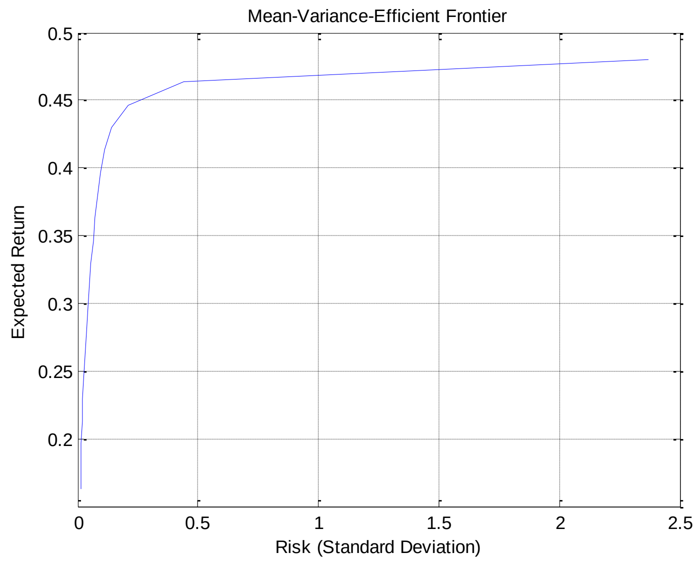

Figure 4 – Efficiency Frontier for the P2P Loan Universe

The figure depicts the efficiency frontier constructed based on the loan classes identified through classification and regression tree analysis. The expected return and risk is the mean return and maturity default probability of loans comprising the loan class portfolios on the frontier.

Since the optimal portfolio does not consist of investments in all loan classes, judging from the lender perspective, it is interesting to know which funding criteria should have been applied in order to select 49 out of 271 loan groups for forming the optimal portfolio and maximizing the Sharpe Ratio. Table 1 in Appendix B reports the weight of each loan class in the optimal portfolio as well as loan and borrower specific criteria that were used to form those classes. The breakdown of the portfolio, however, provides the grounds for the suggestion that there is no preferred either borrower or loan specific characteristic for achieving the optimal return on peer-to-peer lending. These findings do not support evidence from such papers as Meyer (2007) that show that investments in low rating P2P loans should be avoided. Instead, this paper suggests that if these investments are balanced with high quality loans (for example, the optimal portfolio consists of at least 22.72% of AA-A loans), the combined portfolio could not only be yielding positive return, but also deliver the highest remuneration per unit of undertaken risk. In other words, the diversification among loans with several contrary loan characteristics such as in loan classes 3 and 4 with respect to revolving balance utilization or 81 and 82 with respect to revolving balance value (all depicted in Table 1 (Appendix B)) is the suggested key to formation of the most optimal portfolio on Prosper peer-to-peer lending marketplace.

All in all, the empirical evidence presented in this section allows rejecting our main hypothesis regarding the relative unattractiveness of the peer-to-peer lending to stock market investment, if investors strive for an optimal portfolio construction during the whole sample period. It is also worth recalling from the methodology that this conclusion could be made based on the assumption that investors have all the universe of P2P loans available for the investment, i.e. they are not limited by capital constraints and all loan types are available on the market for optimal loan portfolio construction.

5.4 Stage IV: Adjustment to Historical Monthly Loan Availability

In order to relax the first assumption made when comparing the effectiveness of investment into the peer-to-peer lending and stock markets, the sequence of monthly Sharpe Ratios from July 2007 to March 2012 for both markets has been constructed to take the limited loan class variety from month to month on the peer-to-peer lending platform.

Based on the volatility of S&P 500 observed from Figure 5, the time series of stock returns could be separated into two periods – preface of the financial crisis and the period of the latter itself. Despite the higher riskiness of the average investment during the financial crisis and its aftermath, the average Sharpe Ratio of this period (2.55) is higher than the one of the preface period (-1.14). This suggests that investors that were able to time their investments wisely were also able to achieve sufficiently high remuneration per unit of risk despite the financial crisis.

The Sharpe Ratio of P2P lending market has been relatively more stable, following a positive trend throughout the sample period. Overall, the average P2P Sharpe Ratio has been 0.62. The positive dynamics of the indicator could be explained by two following factors:

- The lack of diversification opportunities on Prosper marketplace for forming the highly optimal loan portfolio in the first months and the increasing variety of loan classes over the time, thus, improving investors' diversification opportunities;
- Lender ability to learn from their mistakes and consequently improve their funding decisions, choosing loans with higher expected returns per unit of undertaken risk.

These two suggestions are also supported by the increasing curvature of the P2P market efficiency frontier from month-to-month, which is demonstrated in Figure 1 of Appendix B, and confirmed by similar findings of Freedman & Jin (2011). Besides, the P2P market could have also benefited from any spillover effects of stock market downturn, which could have pushed individuals to search alternative funding sources.

However, two two-sided tests results undermine the superiority of peer-to-peer lending relatively to stock market with respect to the ability to yield more efficient returns observed on Stage III. There is only a 14.60% (t-stat = -1.06) probability with respect to the null hypothesis regarding P2P Sharpe Ratio greatness over S&P 500 to be true. These findings suggest that the variety of loans existing on the P2P platform every month is currently not sufficient to achieve maximum diversification effect in order to compete in terms of risk-return remuneration with the stock market. Therefore, considering the current monthly supply of different loans on the Prosper marketplace, the evidence presented on this stage does not give grounds for rejecting the main hypothesis regarding the stock market superiority over P2P loan market in terms of risk remuneration.

It is also worth noting that the high Sharpe Ratio levels by S&P 500 during the financial crisis period could have been achieved only by highly financially intelligent investors due to importance of the market timing of monthly transactions. As a matter of fact, if one scrutinizes the preface period of the financial crisis, characterized by lower stock market volatility, the probability of the P2P Sharpe Ratio being higher than the S&P's Sharpe Ratio increases up to 98.70 (t-stat = 2.42). Thus, results suggests that in such cases where there are smaller opportunities of gaining high return via intelligent market timing, the Sharpe Ratio of P2P lending market would be at least as high as the one of the stock market.

To sum up, as it was expected, the attractiveness of P2P lending market decreases once the assumption of the widest loan class selection and long investment horizon is relaxed. The peer-to-peer lending market in its current condition could not compete with stock market risk-return ratios on short time intervals, due to higher persistence of P2P loan market returns and limited diversification benefits. The combination of Stage III and IV results seems to suggest that the peer-to-peer lending market is more attractive for investors with longer investment horizons or with lower financial literacy essential for gaining from the stock market volatility.

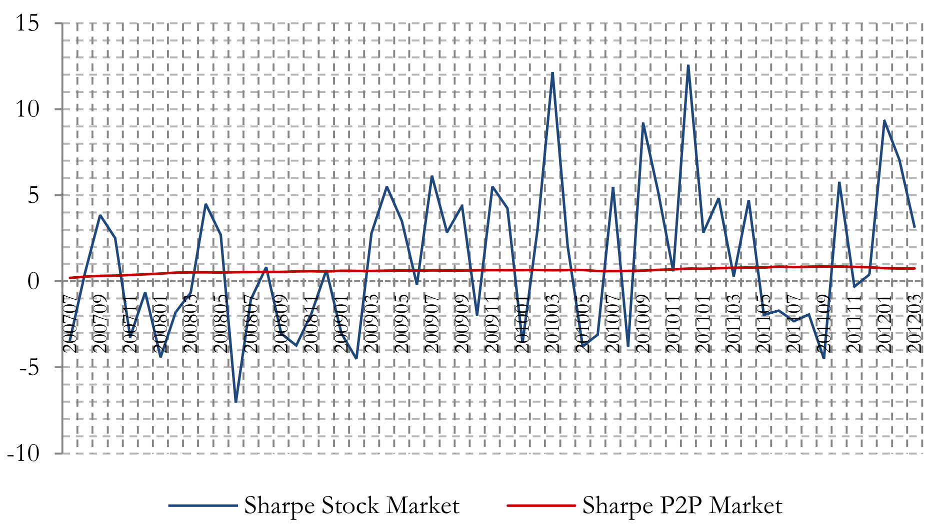

Figure 5 – Sharpe Ratios for P2P Loan and Stock Markets

The figure presents the series of Sharpe Ratios for P2P loan and S&P 500 for full months in the sample period – from July 2007 to March 2012. The ratios for P2P loan market have been constructed based on the risk and return of the most optimal portfolio on the efficiency frontier for each month. The ratios for S&P 500 have been constructed based on the monthly return and the market volatility within the month.

5.5 Stage V: Adjustment to Lender Irrationality

The last stage relaxes the second underlying assumption behind Stage III results regarding the inefficiency of interest setting by lenders. This stage uses actual interest rates that were originally set by lenders instead of the maximum ones in re-applying the methodology of Stages II-IV.

The average expected yield on the peer-to-peer lending market is 23.81% against 24.43% with the efficient interest rate setting mechanism. As it was expected, the lower average return is also linked with lower Sharpe Ratios and even lower attractiveness of the P2P loan market relatively to the stock market if lenders are inefficient in setting correct borrowing rates. Table 1 (Appendix C) exhibits the breakdown of expected returns and respective Sharpe Ratios across loan grades. The splitting factors in classification and regression tree, which is presented in Figure 1 (Appendix C), for determination of the most attractive P2P loan market Sharpe Ratio does not differ from the effective interest rate setting case. As it was expected, the optimal return that could be reached by combining the greatest variety of loan classes for maximizing diversification benefits turned out to be lower – 42.46% instead of 42.64% if lenders were rational. The most optimal Sharpe Ratio remained roughly around 3.20, which is still sufficiently higher than the one of the stock market. This confirms the findings of Stage III and suggests that even with inefficient interest rate setting the lenders can achieve higher risk remuneration than that of the stock market if they follow portfolio optimization strategy. Figure 6 depicts the efficiency frontier of expected yields based on inefficiently set interest rates.

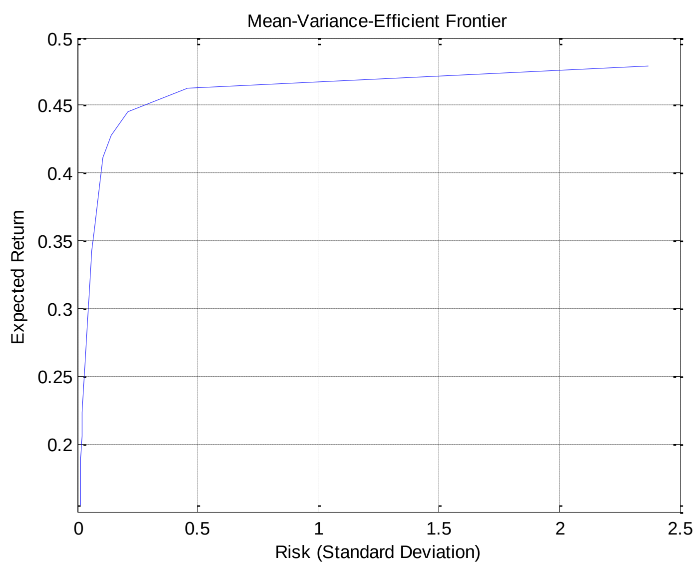

Figure 6 – Efficiency Frontier for the P2P Loan Universe with Inefficient Borrowing Rates

The figure depicts the efficiency frontier constructed based on the loan classes identified through classification and regression tree analysis using the inefficiently set borrowing rates. The expected return and risk is the mean return and maturity default probability of loans comprising the loan class portfolios on the frontier.

Looking at the historical month-to-month performance of P2P lending market and S&P 500 with respect to Sharpe Ratios, the identical relationship could be established as in the result discussion on Stage IV, which confirms the initially presented evidence. For the overall sample period there is only 13.95% (t-stat = -1.09) probability that P2P Sharpe Ratio is on average higher than the one of S&P 500, while for the pre-crisis period the same probability increases to 98.58% (t-stat: =2.37). Both of probabilities are lower than in case of maximized interest rates – 14.60% and 98.70% respectively.

Additionally, Figure 2 (Appendix C) compares the development of optimal Sharpe Ratio series built on efficient and inefficient interest rates and suggests that two series are converging over time. This supports further the suggestion presented in this paper and in Freedman & Jin (2011) that lenders learn from their mistakes, in this case especially with respect to efficient interest rate setting. Despite this finding, investments in peer-to-peer loans still remain relatively unattractive for investors with short investment horizons as well as for those with higher financial literacy. At the same time, the relative attractiveness of peer-to-peer loans as an investment alternative, indeed, decreases if lenders are imprecise in determining the rate that could have been charged from the borrower.

To sum up, Stage V provides evidence that initially presented results are robust with respect to lender irrationality in borrowing rate setting.

5.6 Result Future Implications

The comparison of the risk remuneration on the peer-to-peer lending market has shown that the former market can be attractive for investors with longer investment horizons or for investors with low financial literacy unable to leverage on the stock market volatility. Since Prosper and its rivals promote P2P lending platforms to potential investors as an opportunity for every Average Joe, lending money directly to borrowers without a middleman or initial costs, these platforms have in fact attracted on average a financially illiterate pool of lenders. Considering these facts, one can infer that Prosper targets the right group of people and offers them the opportunity to achieve even higher remuneration for the risk than on the stock market. Besides the private borrowers that have grounds to consider P2P loan marketplaces as improvement in financial markets increasing capital availability, based on the paper results, the private lenders can now question necessity of banks and fund management companies, since functions of both of these institutions could be replicated by lenders themselves on peer-to-peer lending platforms. In order to secure peer-to-peer loan market attractiveness in the long run across all investment horizons and investor types, the results of the thesis give grounds for suggesting two major areas of further improvement of P2P lending platform from investor point of view – increasing the ability of investors to form the most optimal portfolio and increasing the expected payback from every loan.

Judging upon the change in the curvature of efficiency frontiers presented on Stage IV of the analysis and depicted in Figure 1 in Appendix B, the positive effect of a greater selection of available loan classes on the achievable optimal Sharpe Ratio have been spotted. Therefore, if peer-to-peer lending markets, and Prosper in particular, would like to improve the relative risk remuneration for the lenders, the possibility to construct more optimal portfolios by having a selection of loans classes as wide as possible should be secured. Yet, the optimal portfolio allocation across loan classes changes over time depending on peer-to-peer lending market conditions and external macroeconomic factors. As a consequence, in order to allow lenders to adjust their portfolios for these changes, a liquid secondary market for P2P loans should be established. This will also contribute to an increasing attractiveness of the peer-to-peer lending market as an alternative investment class for shorter horizons. Furthermore, directing resources for Prosper development to raising financial literacy of an average lender is not recommended, because it can lead to the latter shifting away from the P2P market to the stock market investment in search for higher gains in shorter periods of time. Instead, a personal portfolio analysis tool on Prosper, with particular suggestions of loans currently in funding stage that would be suitable for a lender in order to achieve higher remuneration for the risk of his/her overall loan portfolio, could be more relevant. Finally, as Figure 2 in Appendix C suggests, the lenders seem to have a steep learning curve with respect to not only funding loans with better risk-return ratio, but also setting interest rates more precisely. In light of this finding it is not clear whether the shift of Prosper to intermediated interest rate setting model will result in more efficient portfolio creation on the platform.

Platforms should also engage so as to increase the average payoff of P2P loans, specifically in two complementary ways. First, the average creditworthiness of borrowers of funded loans could be increased. This could be achieved either by introducing stricter platform policies on accepting loan applications or by increasing incentives for information disclosure by borrowers or by further increasing incentives for groups to act as mini-middlemen and to pre-screen loan applicants among their group members. It is arguable whether the strategy aimed at increasing borrowers' average creditworthiness would result in breaking the image of a no-middleman market. However, by some researchers, such as Freedman & Jin (2011) and Klafft (2008), this shift is seen as an inevitable process of banking re-intermediation and development of peer-to-peer lending markets. Consequently, it could lead to direct competition of P2P platforms with banks in terms of capital provision, as suggested by Freedman & Jin (2011), Dhand et al. (2008) and Ryan et al. (2007); and it could also lead to a direct competition with stock markets in terms of remuneration per unit of risk associated with the investment. Second, even if the borrower defaulted, the recovery rate of these loans could be improved. In order to achieve this, the legal obligations of peer-to-peer loan contracts could be raised as well as higher cooperation with loan originating institutions and debt collection agencies could be developed.

6 Conclusions

6.1 Concluding Thoughts

This paper contributes to the debate on the under-researched lenders' perspective on the emerging concept of peer-to-peer lending. Specifically, it fills the gap in the existing literature by performing the relative assessment of investment risk remuneration on the peer-to-peer lending market and the traditional risky investment alternatives. For this purpose a five-stage methodology has been employed.

The first two stages have been concerned with the calculation of two main components for the assessment of risk remuneration – risk and return. Stage I applied the Cox proportionate hazard model for default probability modeling with results supporting evidence provided by Duarte et al. (2010) that on the peer-to-peer loan market both hard and soft information sources are at least equally important in forecasting loan performance. Stage II constructed individual loan expected returns, which in combination with calculated probability of default, on average leads to a Sharpe ratio of 0.587 instead of 1.25 for the S&P 500. Although this initial evidence shows that an investment into an average loan on the Prosper market place is less attractive comparing to the stock market, these findings suggest that lenders do not lose money per se, which goes in contrast with the findings of Klafft (2008), Freedman & Jin (2011) and Ceyhan et al. (2011).

If investments in P2P loans are not on average loss making, Stage III raises a question whether peer-to-peer loans can at all beat the stock market in the best-case scenario. The latter is characterized with the maximum diversification opportunities on the P2P loan market and with lenders both being efficient in interest rate setting and having a long-term investment horizon following risk-return optimization strategy. The classification of loans with similar characteristics concluded that the loan grade, which is a commonly used variable for grouping loans with respect to expected default rates and expected returns, is not the primary splitting variable. Instead, soft factors such as group ranking and loan purpose are more useful in classifying P2P loans with similar characteristics. The most optimal return achievable on the efficiency frontier based on P2P loan classes is equal to 42.64%, with a Sharpe Ratio notably higher than the one of S&P 500 – 3.20 instead of 1.25 – suggesting relatively higher risk remuneration of P2P loan investments. The analysis of the most optimal portfolio composition does not suggest any particular loan characteristic signaling investment opportunities leading to the most optimal risk-return ratio. Instead, the key to the latter is in portfolio diversification across all loan classes.

The last two stages are concerned with adjusting the best-case scenario to the actual level of the Prosper marketplace development. First, Stage IV adjusts preceding stage results to the actual loan availability on the marketplace, which limits diversification benefits for lenders. Indeed, month-to-month data shows that due to higher P2P loan return persistency and high stock market volatility, the social lending market loses its attractiveness. This suggests that the stock market is more attractive in times with higher volatility particularly for investors who have short enough investment horizon and high enough financial literacy to gain on the monthly volatility of stock returns. Despite this finding, the paper provides evidence of increasing optimal Sharpe Ratio achievable on the peer-to-peer lending market over the sample period, which suggests the positive effect of wider loan class availability and of lender learning process to consequently fund loans with better risk remuneration. Apart from the findings of Freedman & Jin (2011), the lender learning effect is also supported by results of Stage V. In addition to proving the robustness of previous conclusions to lender inefficiency in setting interest rates, Stage V shows that lenders learn not only to select proper loans, but also set interest rates closer to the optimal level.

To sum up, the analysis concludes that the Prosper marketplace as a representation of the peer-to-peer lending market can offer a relatively more attractive investment risk remuneration than the stock market, under the conditions of long investment horizon, low financial literacy of lenders and maximum diversification opportunities. This thesis also suggests that, in order to increase the attractiveness of the market in the long run, social lending platforms need to organize a liquid secondary market for P2P loans, to put more emphasis on filtering out low quality borrowers and to increase cooperation with debt collection agencies for increasing the average recovery rates.

6.2 Suggestion for Further Research

Last, further research can be addressed on the social lending lenders' perspective. Indeed, with the pioneering peer-to-peer lending platform started in 2005, the social lending market could still be considered as a relatively young one, and currently undergoing the growth period. As such, a longer time series of market performance during different life cycle stages and different external macroeconomic conditions would have made the findings of this research more robust. Moreover, the methodology part concerning efficiency frontier construction relies on the assumption of zero correlation between loan classes, which has been accepted in the existing literature, yet is rather strong. The assumption introduces an upward effect on obtained results and could be eliminated in case of availability of longer time series of identical loans on the secondary P2P loan market. Further research in this area can always perform a similar analysis on the enlarged dataset to test persistence of the results of this study over time.

Further, this paper has chosen the Prosper peer-to-peer lending platform as the object of the research due to the non-intermediated nature of loan pricing and funding processes. However, there is no homogeneity in the operational underlying models of Prosper rivals (see Section 2.1 and specifically Figure 1). Due to limitations in accessibility of data and the research time frame, the analysis has been limited to one platform. Despite Prosper being among the most successful P2P lending platforms, future research papers could apply the outlined methodology to other P2P marketplaces and test whether conclusions also hold in different geographical markets and with different P2P platform operation models.

Thirdly, due to limitation in access to the data, this research has incorporated a relatively limited number of soft information sources in estimation of default probabilities and return determination. Future research applying the methodology of this paper could incorporate extra intangible factors such as gender, age, race, beauty of the borrower in papers of Herzenstein et al. (2008, 2011), Ravina (2008), Pope & Sydnor (2011) and Duarte et al. (2010) in order to achieve higher precision of default probability and return estimation models.

Finally, just as the other researches on the lender perspective of peer-to-peer lending markets, the study is implicitly based on the major assumption that all loans that have been funded should have been funded by investors. Since the data on failed loan applications was not available for this study, it would be worth analyzing the overall diversity of available loan applications that could have been funded in order to achieve higher return efficiency on the peer-to-peer lending market. Consequently, the efficiency of current funding choices of lenders with respect to achieving higher remuneration for the risk could also be investigated.

REFERENCES

Akerlof, G. A. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, *84*(3), 488-500. JSTOR.

Arora, N., Bohn, J. R., & Zhu, F. (2005). Reduced Form vs . Structural Models of Credit Risk : A Case Study of Three Models. *Managing*, *3*(4), 43-67. Citeseer.

Ashta, A., & Assadi, D. (2009). Do Social Cause and Social Technology Meet ? Impact of Web 2 . 0 Technologies on peer-to-peer lending transactions. *Forum American Bar Association*, *23*(June 2008), 1-35.

Andreeva, G., Lin, M. & Ansell, J. (2007). Merton models or credit scoring : modelling default of a small business. University of Edinburgh Management School, Edinburgh

Bachmann, A., Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., & Tiburtius, P. (2011). Journal of Internet Banking and Commerce. *Journal of Internet Banking and Commerce*, *16*(2).

Barasinska, N., Schäfer, D. (2010). Does Gender Affect Funding Success at the Peer-to-Peer Credit Markets? Evidence from the Largest German Lending Platform. *Working paper*.

Berger, S. C., & Gleisner, F. (2009). Emergence of Financial Intermediaries in Electronic Markets : The Case of Online P2P Lending. *Online*, *2*(1), 39-65.

Böhme, R., Pötzsch, S. (2010). The Role of Soft Information in Trust Building: Evidence from Online Social Lending. Proc. of International Conference on Trust and Trustworthy Computing (TRUST), LNCS 6101, Springer-Verlag, Berlin Heidelberg, pp. 381–395.

Caldieraro, F., Jr, M. C., Shulman, J. D., & Zhang, J. (2011). Is Silence Golden ? – How Non-Verifiable Information Influences Funding Outcomes On Peer-to-Peer Lending Platforms. Marketing Department at Michael G. Foster School of Business, University of Washington, (April), 1-39.

Çetin, U., Jarrow, R., Protter, P., Yildirim, Y. (2004). Modeling Credit Risk with Partial Information, Annals of Applied Probability, Vol. 14, Issue 3

Ceyhan, S. (2011). Dynamics of Bidding in a P2P Lending Service : Effects of Herding and Predicting Loan Success. *Analysis*, 547-556. ACM.

Chava, S. & Jarrow, R.A. (2004). Bankruptcy Prediction With Industry Effects, Working Paper, August 2004

Collier, B., Hampshire, R., Heinz, H. J., & College, I. I. I. (2010). Sending Mixed Signals : Multilevel Reputation Effects in Peer-to-Peer Lending Markets.*Forbes*, 197-206. ACM Press.

Crocker, J., Major, B., and Steele, C. (1998). Social stigma. in D.T. Gilbert, S.T. Fiske and G.Lindzey (eds), The Handbook of Social Psychology (Vols 1 and 2), 504-553. McGraw-Hill.

Dhand, H., Mehn, G., Dickens, D., Patel, A., Lakra, D., & McGrath, A. (2008). Internet Based Social Lending. *Money*, *2*, 109-114.

Duarte, J., Siegel, S., Young, L. (2010). Trust and Credit. Working Paper. 1st round review, *Journal of Finance*.

Everett, C. R. (2010). Group membership , relationship banking and loan default risk : the case of online social lending. *Group*, *2011*(765), 39.

Freedman, S., Jin, G.Z. (2011). Learning by Doing with Asymmetric Information: Evidence from Prosper.com. National Bureau of Economic Research, Inc, *NBER Working Papers*: 16855.

Freedman, S., Jin, G.Z. (2008). Do Social Networks Solve Information Problems for Peer-to-Peer Lending? Evidence from Prosper.Com. *NET Institute Working Paper No. 08-43*.

Galloway, I. J. (2009). Peer-to-peer lending and community development finance. *Community Development Investment Center Working Paper*, *21*(3), 18-39. Federal Reserve Bank of San Francisco.

Greiner, M. E., & Wang, H. (2009). The Role of Social Capital in People-to-People Lending Marketplaces. *ICIS 2009 Proceedings*, 18. Association for Information Systems.

Greiner, M., & Wang, H. (2007). Building Consumer-to-Consumer Trust in e-Finance Marketplaces. *AMCIS 2007 Proceedings* (Vol. 211, p. 11). Association for Information Systems.

Heng, S., Meyer, T., Stobbe, A. (2007). Be a driver, not a passenger: Implications of Web 2.0 for financial institutions. *Deutsche Bank Research*.

Herrero-Lopez, S. (2009). Social Interactions in P2P Lending. *Proceedings of the 3rd Workshop on Social Network Mining and Analysis SNAKDD 09*, *09*, 1-8. ACM Press.

Herzenstein, M., Dholakia, U. M., & Andrews, R. L. (2010). Strategic Herding Behavior in Peer-to-Peer Loan Auctions. *Journal of Interactive Marketing*, *In Press*(713), 1-32. Direct Marketing Educational Foundation, Inc.

Herzenstein, M., Sonenshein, S., & Dholakia, U. M. (2011). Tell Me a Good Story and I May Lend You Money: The Role of Narratives in Peer-to-Peer Lending Decisions. *Journal of Marketing Research*, *48*(SPL), S138-S149. Rice University.

Hildebrand, T., Puri, M., & Rocholl, J. (2010). Skin in the Game : Evidence from the Online Social Lending Market. *Group*, (October).

Hothron, T., Hornik, K., and Adam, Z. (2006). Unbiased recursive partitioning: a conditional inference framework. Journal of computational and Graphical Statistics, 15 651-654.

Iyer, R., Ijaz, A., Erzo, K., & Kelly, F. P. L. (2011). Inferring Asset Quality : Determining Borrower Creditworthiness in Peer-to-Peer Lending Markets.*Context*, (June).

Klafft, M. (2008). Online Peer-to-Peer Lending : A Lenders ' Perspective. (H. R. Arabnia & A. Baharami, Eds.)*International Journal*, *SSRN*.

Klein, T. (2008). Performance in Online Lending Platforms Diplomarbeit.*Online*, (September 2007), 1-42.

Krumme, K. A., & Herrero, S. (2009). Lending Behavior and Community Structure in an Online Peer-to-Peer Economic Network. *2009 International Conference on Computational Science and Engineering*, *4*, 613-618.

Kumar, S. (2007). Bank of One: Empirical Analysis of Peer-to-Peer Financial Marketplaces. *Americas Conference on Information Systems* (p. 9). Association for Information Systems.

Lin, M., Prabhala, N. R., & Viswanathan, S. (2009). Judging Borrowers By The Company They Keep : Social Networks and Adverse Selection in Online Peer-to-Peer Lending. *Working paper*.

Meyer,T. (2009). Innovations in P2P lending may put computers over people: Welcome to the machine. *Deutsche Bank Research*

Meyer,T. (2007). The power of people: Online P2P lending nibbles at banks' loan business. *Deutsche Bank Research*.

Miller, S. (2011). Risk Factors for Consumer Loan Default: A Censored Quantile Regression Analysis. *Working Paper*.

Pauly, Mark V. (1974): Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection, Quarterly Journal of Economics, 88: 44-62.

Pope, D. G. & Sydnor, J. R. (2008). What's in a Picture ? Evidence of Discrimination from Prosper . com. *Journal of Human Resources*, *46*(1), 53-92

Qiu, J., Xu, Y., Chen, D., & Zhang, G. (2011). The effects of social capital In Chinese online P2P lending market. *2011 International Conference on Innovation and Information Management* (Vol. Chengdu).

Qiu, J., Xu, Y., Chen, D., & Zhang, G. (2010). The roles of social capital in online P2P lending markets under different cultures: a comparison of china and America. *The Tenth International Conference on Electronic Business*.

Ravina, E. (2008). Love & Loans The Effect of Beauty and Personal Characteristics in Credit Markets. *Symposium A Quarterly Journal In Modern Foreign Literatures*, *972801*(2), 560-573.

Rumiany, D. (2007). Internet Bidding for Microcredit: making it work in the developed world, conceiving it for the developing world. *Development Gateway March*.

Ryan, B. J., Reuk, K., & Wang, C. (2007). Determinants of Loan Fundability in the Prosper . com Marketplace To Fund Or Not To Fund. *Business*, *2011*(8/15), 1-23.

Sharpe, W. F. (1966). Mutual Fund Performance. *The Journal of Business*,*39*(1), 119-138. JSTOR.

Shumway, T. 2001, "Forecasting Bankruptcy More Accurately: A Simple Hazard Model", Journal of Business, January 2001

Singh, H., Gopal, R., Li, X. (2009). Risk and Return of Investments in Online Peer-to-Peer Lending. *Proceedings of the 19th Workshop on Information Technologies and Systems (WITS 2009).*

Slavin, B. (2007). Peer-to-Peer Lending - An Industry Insight. *Online*. Retrieved from http://www.bradslavin.com/wp-content/uploads/2007/06/peer-to-peer-lending.pdf

Srinivasan, R. & Sriram, M.S. (2003). Microfinance: An Introduction", Round Table, *IIMB Management Review*, June 2003

Strasser, H., and Weber, C. (1999). On the asymptotic theory of permutation statistics. *Mathematical methods of statistics* 8 220-250.

Timofeev, R. (2004). Classification and Regression Trees (CART) Theory and Applications. *Master Thesis*, Humboldt University, Berlin, 1-40.

United States Government Accountability Office. (2011). Person-To-Person Lending - New Regulatory Challenges Could Emerge as the Industry Grows, *Report to Congressional Committees*, July 2011

Weiß, G. N., Pelger, K., & Horsch, A. (2010). Mitigating Adverse Selection in P2P Lending - Empirical Evidence from Prosper.com. *thejournalofrisk.com*. *Working paper*.

Whalen, G. (1989). A Proportional Hazards Model for Bank Failure: An Examination of Its Usefulness as an Early Warning Tool, Federal Reserve Bank of Cleveland, *Economic Review*, Quarter 1

Wu, J., & Xu, Y. (2011). A Decision Support System for Borrower's Loan in P2P Lending. *Journal of Computers*, *6*(6), 1183-1190.

ADDITIONAL REFERENCES

French Data Library. (2012). *Online*. http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html; Accessed on April 19, 2012

LendingClub Loan Performance Page. (2012). *Online*. https://www.lendingclub.com/info/statistics-performance.action; Accessed on April 19, 2012

Prosper Website - About Us. (2012). *Online*. http://www.prosper.com/about/; Accessed on April 19, 2012

Prosper Website - Media Room. (2012). *Online*. http://www.prosper.com/about/media_press_releases.aspx; Accessed on April 19, 2012

Prosper Website - Prosper Investor Help. (2012). Online. http://www.prosper.com/help/investing.aspx; Accessed on April 19, 2012

Appendix A – Descriptive Statistics

Table 1 – Loan Grade and Financial Information

Loan Grade	N	Amount	Term	Borrower Rate	i(max) – i(actual)	Time Till The Last Payment Made	Not Fully Paid	Description Length	Inquiries	Revolving Balance	Revolving Balance Utilization	DTI	Group Ranking
Total	11752	9787.09	39.278	12.325	1.466	0.798	29.15%	344.504	1.519	15087.664	46.673	12.463	6.633
AA	2278	7240.83	36.643	7.776	2.152	0.917	12.73%	290.897	0.867	11034.043	24.125	10.111	4.320
A	3104	10144.45	38.590	10.936	0.687	0.831	24.65%	337.164	1.197	14585.992	40.551	11.989	5.934
B	2730	9618.66	38.901	12.980	0.696	0.781	31.83%	347.990	1.608	14719.730	52.637	12.894	6.925
C	1883	10383.75	40.232	14.781	1.380	0.730	38.34%	374.331	1.887	17300.204	58.435	13.430	7.902
D	1088	11274.84	43.566	16.224	3.816	0.700	66.36%	378.108	2.205	16406.429	63.470	13.850	8.647
E	441	13587.36	44.109	17.917	2.716	0.658	47.17%	418.737	2.517	20846.576	66.433	15.227	9.490
HR	228	13001.54	41.789	19.174	1.507	0.649	52.19%	388.022	3.118	31118.763	68.397	17.307	10.114

The table presents descriptive statistics for variables that were initially available for the universe of P2P loans and the ones that were constructed based on the latter. The first row exhibits averages for the overall sample for each variable, while subsequent rows present averages for loan grade subsamples. N stands for the number of observations in the sample; Amount – the monetary amount of borrowed money; Term – the term of the loan contract in months; Borrower Rate – the final rate set to be paid by the borrower; (i(max) – i(actual)) – the difference between the maximum interest rate that could have been charged from a borrower judging upon similar loan and borrowers and the actual rate for the particular loan; Time Till The Last Payment Made – share of the overall term passed until the borrower stopped complying with loan obligations; Not Fully Paid – proportion of loans that were not paid out in full by the end of the loan term; Description Length – the symbol length of the loan description written by the borrower to explain the reason behind the loan; Inquiries – the number of inquiries to the credit agency regarding borrower credit status and history during the last 6 months; Revolving Balance – the monetary amount of revolving credit balance; Revolving Balance Utilization – the percentage of available revolving credit that is utilized at the time the loan is originated; DTI – total debt to income ratio of the borrower at the time of loan origination; Group Ranking – the ranking of the group that the borrower is a member of on the P2P lending platform.

Table 2 – Purpose of the Loan and Financial Information

Purpose	N	Amount	Term	Loan Grade	Borrower Rate	i(max) - i(actual)	Time Till The Last Payment Made	Not fully paid	Description Length	Inquiries	Revolving Balance	Revolving Balance Utilization	DTI	Group Ranking
Total	11752	9787.09	39.28	5.12	12.33	1.47	0.80	0.29	344.50	1.52	15087.66	46.67	12.46	6.63
Debt Consol.	5044	11023.54	39.73	4.91	12.81	1.50	0.79	0.30	329.72	1.43	15399.53	53.50	13.95	6.88
House Improv.	850	10147.74	39.92	5.47	11.58	1.48	0.81	0.28	292.71	1.82	13785.00	34.13	10.03	6.23
Business	719	12801.95	39.81	4.57	13.48	1.45	0.67	0.47	630.28	1.89	22977.11	40.25	10.91	7.22
Card Refin.	1410	9832.18	37.92	5.25	11.90	1.48	0.86	0.21	366.3	1.33	21288.59	53.67	13.79	6.43
Education	276	6644.56	36.52	5.14	12.10	1.35	0.82	0.29	427.51	2.08	9780.70	37.40	10.74	6.54
Car	387	6548.64	41.09	5.83	10.69	1.54	0.88	0.19	275.14	1.36	11002.97	34.63	10.04	5.81
House	161	10782.61	39.28	5.19	12.21	1.33	0.82	0.25	442.73	2.57	8785.75	31.67	9.54	6.55
Major Purch.	631	7635.18	39.31	5.55	11.38	1.49	0.83	0.23	269.32	1.35	8478.97	33.54	9.91	6.15
Medicine	221	7377.72	39.69	5.13	12.24	1.51	0.75	0.34	320.67	1.58	15031.49	43.19	11.32	6.61
Relocation	174	6644.83	38.90	5.30	11.91	1.44	0.77	0.31	309.41	1.74	7559.48	41.05	10.59	6.41
Vacation	97	4940.72	38.72	5.68	11.12	1.52	0.76	0.32	187.03	1.27	7139.44	37.93	10.83	6.01
Wedding	278	9001.71	38.76	5.25	12.12	1.38	0.84	0.22	327.62	1.46	9088.43	38.73	10.80	6.52
Renew. Energy	28	9041.07	41.14	5.75	11.14	1.13	0.66	0.39	561.54	0.93	22920.68	43.19	12.82	6.07
Other	1476	7302.20	38.50	5.27	12.02	1.35	0.79	0.31	305.96	1.55	12980.84	42.49	11.45	6.49

The table presents descriptive statistics for variables that were initially available for the universe of P2P loans and the ones that were constructed based on the latter. The first row exhibits averages for the overall sample for each variable, while subsequent rows present averages for loan purpose subsamples. N stands for the number of observations in the sample; Amount – the monetary amount of borrowed money; Term – the term of the loan contract in months; Borrower Rate – the final rate set to be paid by the borrower; Loan Grade – the grade from AA (1) to HR (7) assigned to the loan by P2P lending platform based on the credit score of a borrower; (i(max) – i(actual)) – the difference between the maximum interest rate that could have been charged from a borrower judging upon similar loan and borrowers and the actual rate for the particular loan; Time Till The Last Payment Made – share of the overall term passed until the borrower stopped complying with loan obligations; Not Fully Paid – proportion of loans that were not paid out in full by the end of the loan term; Description Length – the symbol length of the loan description written by the borrower to explain the reason behind the loan; Inquiries – the number of inquiries to the credit agency regarding borrower credit status and history during the last 6 months; Revolving Balance – the monetary amount of revolving credit balance; Revolving Balance Utilization – the percentage of available revolving credit that is utilized at the time the loan is originated; DTI – total debt to income ratio of the borrower at the time of loan origination; Group Ranking – the ranking of the group that the borrower is a member of on the P2P lending platform.

Table 3 – Loan Grade and Loan Purpose

Loan Grade	Debt Consol.	House Improv	Business	Card Refin.	Education	Car	House	Major Purch.	Medicine	Relocation	Vacation	Wedding	Renew. Energy	Other
Total	0.429	0.072	0.061	0.120	0.023	0.033	0.014	0.054	0.019	0.015	0.008	0.024	0.002	0.126
AA	0.310	0.108	0.036	0.132	0.024	0.064	0.014	0.087	0.022	0.020	0.013	0.025	0.003	0.142
A	0.420	0.077	0.054	0.121	0.023	0.036	0.016	0.053	0.017	0.015	0.010	0.027	0.004	0.129
B	0.440	0.064	0.056	0.129	0.028	0.026	0.012	0.045	0.019	0.015	0.007	0.022	0.003	0.133
C	0.508	0.052	0.070	0.109	0.016	0.020	0.011	0.052	0.013	0.010	0.004	0.024	0.001	0.110
D	0.492	0.052	0.082	0.108	0.028	0.015	0.013	0.032	0.025	0.016	0.006	0.021	0	0.110
E	0.506	0.061	0.125	0.082	0.020	0.011	0.018	0.029	0.023	0.014	0.002	0.011	0	0.098
HR	0.509	0.039	0.180	0.092	0.022	0	0.013	0.004	0.018	0.013	0	0.022	0.004	0.083

The table presents the shares that loan purpose take in loan grade subsamples. The sum of values in each row is equal to 100%.

Table 4 – Final Loan Repayment Status and Financial Information

Final Status	N	Amount	Term	Loan Grade	Borrower Rate	i(max) – i(actual)	Description Length	Inquiries	Revolving Balance	Revolving Balance Utilization	DTI	Group Ranking
Total	11752	9787.09	39.278	5.116	12.325	1.466	344.504	1.519	15087.664	46.673	12.463	6.633
Fully paid	8326	9439.06	38.361	5.330	11.800	1.454	342.812	1.369	14391.778	43.758	12.052	6.368
Not fully paid	3426	10632.89	41.506	4.596	13.602	1.493	348.617	1.884	16778.835	53.758	13.462	7.276

The table presents the averages for each of initially available and constructed variables across subsamples of fully paid and not fully paid loans. N stands for the number of observations in the sample; Amount – the monetary amount of borrowed money; Term – the term of the loan contract in months; Loan Grade – the grade from AA (1) to HR (7) assigned to the loan by P2P lending platform based on the credit score of a borrower; Borrower Rate – the final rate set to be paid by the borrower; (i(max) – i(actual)) – the difference between the maximum interest rate that could have been charged from a borrower judging upon similar loan and borrowers and the actual rate for the particular loan; Description Length – the symbol length of the loan description written by the borrower to explain the reason behind the loan; Inquiries – the number of inquiries to the credit agency regarding borrower credit status and history during the last 6 months; Revolving Balance – the monetary amount of revolving credit balance; Revolving Balance Utilization – the percentage of available revolving credit that is utilized at the time the loan is originated; DTI – total debt to income ratio of the borrower at the time of loan origination; Group Ranking – the ranking of the group that the borrower is a member of on the P2P lending platform.

Table 5 – Final Loan Repayment Status and Loan Purposes

Final Status	Debt Cons.	House Improv.	Business	Card Refin.	Education	Car	House	Major Purch.	Medicine	Relocation	Vacation	Wedding	Renew. Energy	Other
Total	0.429	0.072	0.061	0.120	0.023	0.033	0.014	0.054	0.019	0.015	0.008	0.024	0.002	0.126
Fully Paid	0.423	0.074	0.046	0.134	0.023	0.038	0.014	0.059	0.017	0.014	0.008	0.026	0.002	0.123
Not fully paid	0.445	0.069	0.099	0.087	0.024	0.022	0.012	0.042	0.022	0.016	0.009	0.018	0.003	0.133

The table presents the shares that loan purpose take in fully paid and not fully paid loan subsamples. The sum of values in each row is equal to 100%.

Appendix B – Optimal P2P Loan Portfolio & P2P Market Efficiency Frontier

Table 1 – Optimal Portfolio Breakdown

Loan Class ID	Weight in the Optimal Portfolio	Choice Criterion									Mean Expected Return
		Criterion 1: Term	Criterion 2: Purpose	Criterion 3: Revolving Balance Utilization	Criterion 4: Group Ranking	Criterion 5: Loan Grade	Criterion 6: Debt-to-Income	Criterion 7: Inquiries Last 6 Months	Criterion 8: Revolving Balance	Criterion 9: i(max) – i(actual)	
1	0.591%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding			AA					33.219%
2	3.304%	60	Business, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy		4-5	AA, A					40.790%
3	3.193%	60	Business, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	<25.65%	6-7	AA, A					42.441%
4	2.229%	60	Business, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=25.65%	6-7	AA, A					44.292%
5	2.390%	60	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	<51.2%	7	B, C, D, E, HR					44.271%
6	0.604%	60	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	<51.2%	10-12	B, C, D, E, HR					48.003%
9	0.109%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	<48.25%	9-11			>=2.5			34.854%
10	0.615%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=48.25%	9-11			>=3.5			40.062%
11	1.168%	60	Car, Major Purchases, Wedding	<40.85%		A					35.576%
12	3.632%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	<63.45% & >=40.85%		A					39.772%
13	3.710%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=63.45%		A					40.864%
14	3.772%	60	Credit Card Refinancing, Car, Major Purchases, Wedding	<44.65%	7-8	B, C, D, E, HR					39.116%
15	2.167%	60	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	<8.95%	8-12	B, C, D, E, HR					45.023%
16	1.236%	60	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=51.2%	7-9	B					46.233%
27	0.052%	36	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding		10-11			>=2.5			32.880%
28	0.349%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=48.25%	10-11			<3.5			38.101%
29	1.492%	60	Debt Consolidation, Credit Card Refinancing, House	<40.85%		A	<12.16%				37.463%
30	3.405%	60	Debt Consolidation, Credit Card Refinancing, House	<40.85%		A	>=12.165				39.161%
31	4.127%	60	Debt Consolidation, House	<44.65%	7-8	B, C, D, E, HR	<15.13%				40.678%
32	3.914%	60	Debt Consolidation, House	<44.65%	7-8	B, C, D, E, HR	>=15.13%				41.965%
33	3.874%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=44.65%	7	B, C, D, E, HR	<12.225%				41.169%
34	3.585%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=44.65%	8	B, C, D, E, HR	<15.62%				42.861%
35	4.244%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	<57.75%	9	B, C, D, E, HR	<10.43%				42.161%
36	2.349%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=57.75%	9	B, C, D, E, HR		>=1.5			45.078%
37	1.742%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	<71.8%	11-12	B, C, D, E, HR					46.006%

38	1.239%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=71.8%	11-12	B, C, D, E, HR					46.804%
39	1.619%	60	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=8.95% & <51.2%	8	B, C, D, E, HR					45.795%
40	1.282%	60	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=8.95% & <51.2%	9-12	B, C, D, E, HR					46.584%
41	1.163%	60	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=51.2%	7-9	C, D, E	<12.93%				46.820%
42	0.889%	60	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=51.2%	7-9	C, D, E	>=12.93%				47.259%
72	0.010%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=27.95% & <52.35%	8			>=2.5			32.170%
75	0.019%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=76.4%	8		<11.005%				32.331%
76	0.112%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=52.35%	8		>=11.005%	>=2.5			35.429%
77	0.078%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=48.25% & <85.35%	9			<3.5			33.514%
78	0,137%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=85.35%	9						36.208%
79	3.773%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=44.65% & <64.35%	7	B, C, D, E, HR	>=12.225%				41.795%
80	3.351%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=64.35%	7	B, C, D, E, HR	>=12.225%				42.623%
81	3.275%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=44.65%	8	B, C, D, E, HR	>=15.62%		<15978		43.395%
82	2.872%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=44.65%	8	B, C, D, E, HR	>=15.62%		>=15978		43.930%
83	2.975%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	<57.75%	9	B, C, D, E, HR	>=10.43%			<2.35%	43.909%
84	3.756%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	<57.75%	9	B, C, D, E, HR	>=10.43%			>=2.35%	43.241%
85	2.671%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=57.75%	9	B, C, D, E, HR		<1.5		<4.495%	44.465%
86	3.682%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=57.75%	9	B, C, D, E, HR		<1.5		>=4.495%	43.609%
87	2.942%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	<73.85%	10	B, C, D, E, HR	<14.015%				44.236%
88	2.437%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	<73.85%	10	B, C, D, E, HR	>=14.015%				45.081%
89	2.078%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=73.85%	10	B, C, D, E, HR	<17.095%				45.505%
90	1.669%	60	Debt Consolidation, Credit Card Refinancing, Car, House, Major Purchases, Wedding	>=73.85%	10	B, C, D, E, HR	>=17.095%				46.124%
149	0.035%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=76%	7		>=11.46%	>=1.5			32.871%
151	0.082%	36	Business, Education, House Improvement, Medicine, Relocation, Vacation, Other, Renewable Energy	>=81.5%	8		>=11.005%	<2.5			33.945%

The table demonstrates the most optimal portfolio breakdown for the peer-to-peer lending market assuming the maximum selection of loan classes. The latter are formed through classification and regression tree analysis. Columns 3-9 in the table represent the splitting criteria in the classification and regression tree analysis for forming the loan classes. Term stands for the overall term of the loan contract in months; Purpose – clustered loan purpose of the loan chosen by the borrower; Revolving Balance Utilization – the percentage of available revolving credit that is utilized at the time the loan is originated; Group Ranking – the ranking, from the highest (1) to the lowest (12), of the group that the borrower is a member of on the P2P lending platform; Loan Grade – the grade from AA (1) to HR (7) assigned to the loan by P2P lending platform based on the credit score of a borrower; Debt-to-Income – total debt to income ratio of the borrower at the time of loan origination; Inquiries Last 6 Months – the number of inquiries to the credit agency regarding borrower credit status and history during the last 6 months; Revolving Balance – the monetary amount of revolving credit balance; (i(max)-i(actual)) – the difference between the maximum interest rate that could have been charged from a borrower judging upon similar loan and borrowers and the actual rate for the particular loan. Column 2 shows the share that the specific loan class accounts for in the most optimal portfolio. The last column shows the expected return of the loan class.

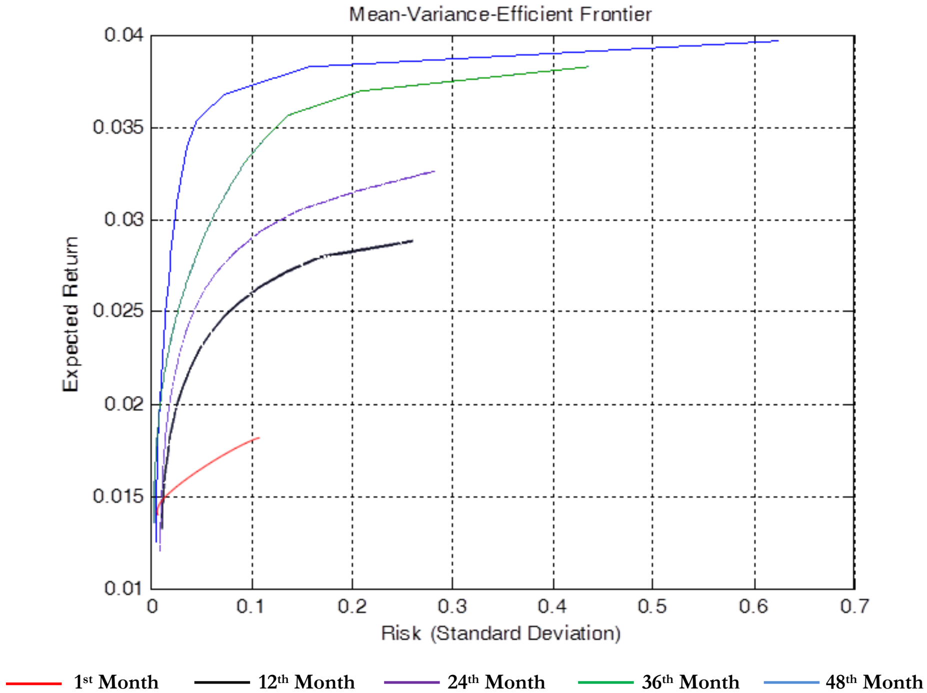

Figure 1 – P2P Efficiency Frontier Dynamics from Month to Month

The figure demonstrates the change dynamics of P2P loan market efficiency frontiers on for every 12th month in the sample period. The frontiers for P2P loan market are constructed based on the available selection of loan classes, which are formed through the classification and regression tree analysis.

Appendix C – Results with Inefficient Interest Rate Setting

Table 1 – P2P Expected Return and related Sharpe Ratio

Loan Grades	N	Average Probability of Defaulting by end of the term	Expected Return	Sharpe Ratio
AA	2278	18.953%	14.520%***	0.696
A	3104	32.979%	21.631%***	0.616
B	2730	43.797%	25.062%***	0.542
C	1883	53.828%	28.190%***	0.499
D	1088	66.733%	31.924%***	0.459
E	441	88.301%	34.552%***	0.376
HR	228	96.361%	35.376%***	0.353

The table presents the distribution of the average probability of defaulting by maturity, expected return and the corresponding Sharpe Ratio across the loan grade subsamples, which consist of N-number of observations. The table is constructed based on the inefficiently set interest rates.

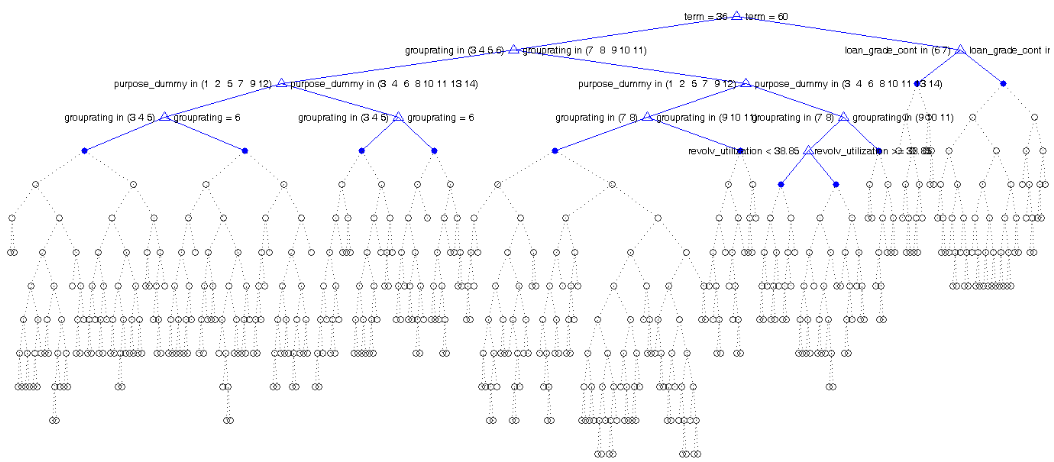

Figure 1 – Classification Tree with Inefficient Interest Rates

The figure depicts the output of the classification and regression tree analysis of the whole sample loan universe. The tree has been constructed based on group loans with similar characteristics of the main variable – expected return. For simplicity only the first splitting nodes and the respective decision criteria are presented. The total number of terminal nodes is 269.

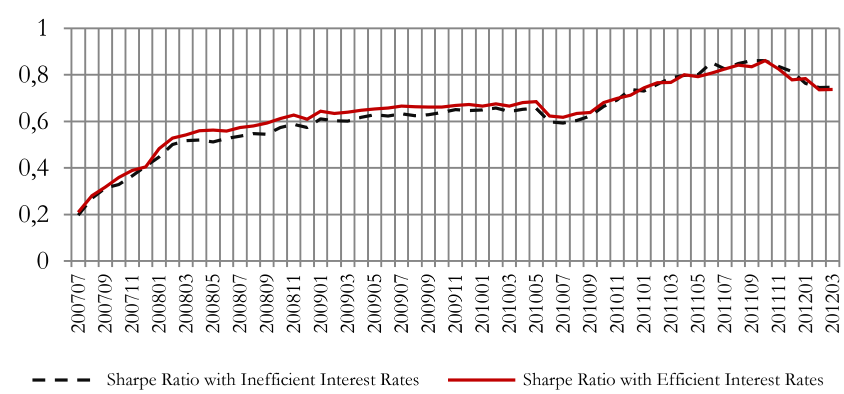

Figure 2 – Sharpe Ratio Comparison with Efficient and Inefficient Interest Rates

The figure presents the historical development of P2P loan market Sharpe Ratios for two scenarios – efficiently and inefficiently set interest rates. Sharpe Ratios are constructed for the most optimal portfolio in each month based on the available selection of loan classes, which are determined through classification and regression tree analysis.